HOMEWORK 1

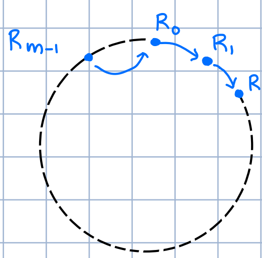
Anvita Bhagavathula

1. Choose an $m \in \{150, 151, \cdots, 5000\}$ and an $a \in \{2, 3, \ldots, m-1\}$ so that the multiplicative congruential generator (MCG) $R_n = (aR_{n-1}) \bmod m$ has full period (no holes). For the seed $R_0 = 1$, generate $R_1, \ldots, R_{m-1}$ and define N_K to be the number of R_j's that are equal to K. More formally, $N_K = \#\{j : R_j = K,\ 1 \le j \le m-1\}$. What should N_K be for each $K = 1, 2, \ldots, m-1$.

↳ Let us choose an $m \in \{150, \ldots, 5000\}$ and an $a \in \{2, 3, \ldots, m-1\}$
↳ using MATLAB to check whether m is prime, we can obtain →

$\Rightarrow$ $m = 173$
$a = 7$ → multiplicative constant

↳ For the seed $R_0 = 0$, let us generate the pseudo-random sequence $R_1, R_2, \ldots, R_{m-1}$ →



$$
\begin{aligned}
R_0 &= 1 \\
R_1 &= 7R_0 \bmod 173 = 7 \bmod 173 = 7 \\
R_2 &= 7R_1 \bmod 173 = 49 \bmod 173 = 49 \\
R_3 &= 7R_2 \bmod 173 = 343 \bmod 173 = 170 \\
R_4 &= 7R_3 \bmod 173 = 1190 \bmod 173 = 152 \\
&\vdots \\
R_{172} &= 7R_{171} \bmod 173 = 693 \bmod 173 = 1
\end{aligned}
$$

↳ We can generate the full sequence $R_1, \ldots, R_{m-1}$ through MATLAB to obtain →

$\therefore$ $\{R_1, \ldots, R_{m-1}\} = \{7, 49, 170, 152, 26, \ldots, 99, 1\}$, where full sequence is given by running attached MATLAB code for question 1.

↳ Now let us determine $N_K = \#\{j : R_j = K,\ 1 \le j \le m-1\}$ for $K = 1, 2, 3, \ldots, m-1$

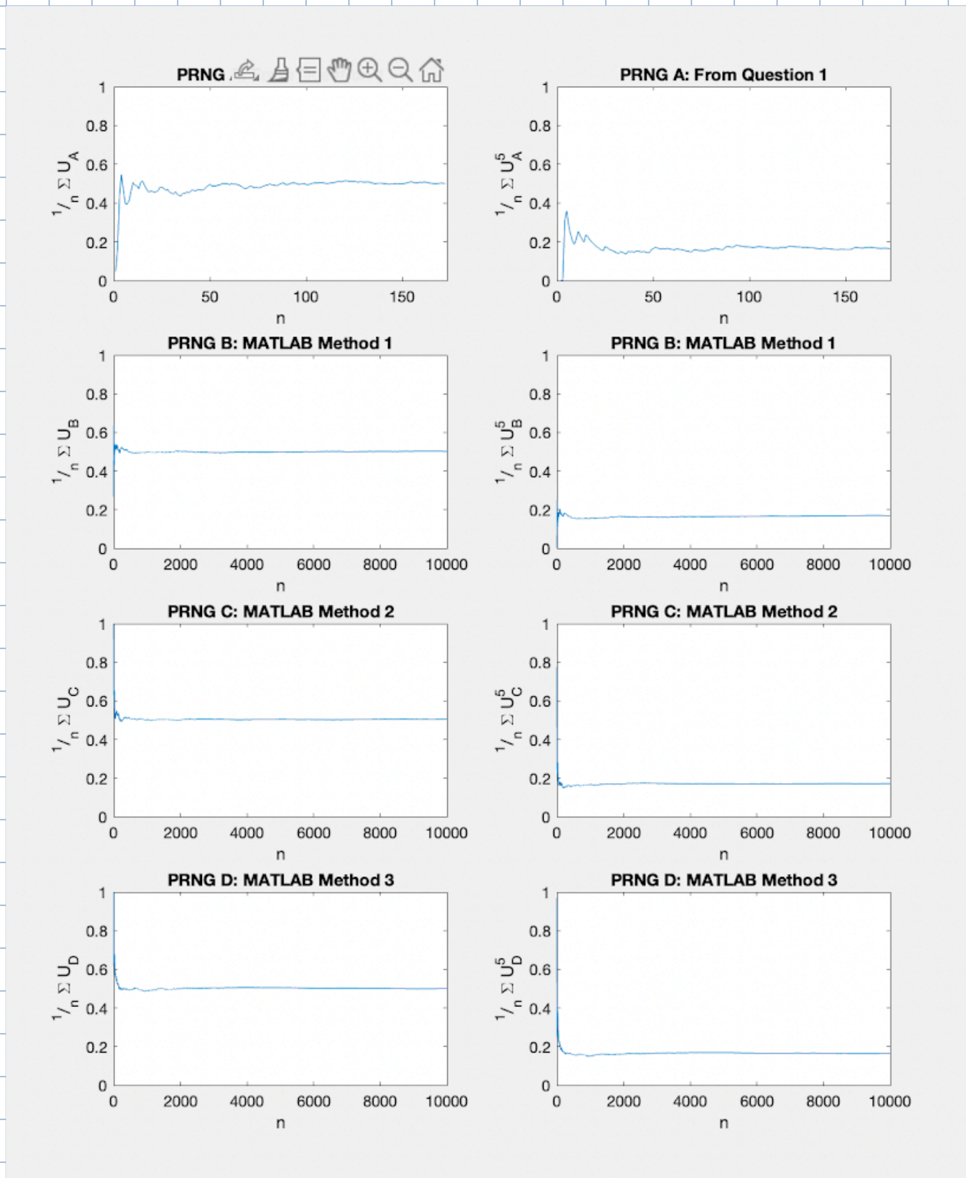
$$
\begin{aligned}
\Rightarrow N_1 &= \#\{j : R_j = 1,\ 1 \le j \le 172\} = 1 \\
N_2 &= \#\{j : R_j = 2,\ 1 \le j \le 172\} = 1 \\
&\vdots \\
N_{m-1} &= \#\{j : R_j = 172,\ 1 \le j \le 172\} = 1
\end{aligned}
$$

$\therefore$ $N_K = \#\{j : R_j = K,\ 1 \le j \le m-1\}$ for each $K = 1, 2, \ldots, m-1$ is given by $\{1, 1, \ldots, 1\}$ where the full set N_K can be found by running attached MATLAB code for question 1.

2. For each of the four PRNGs defined in the question, let $U_1, \ldots, U_{1000}$ be a sequence from the PRNG.
Draw two plots of →

$\frac{1}{n}\sum_{K=1}^{n} U_K$ versus n, for $n = 1, \ldots, 10000$

$\frac{1}{n}\sum_{K=1}^{n} U_K^5$ versus n, for $n = 1, \ldots, 10000$



↳ for $\frac{1}{n}\sum_{K=1}^{n} U_K$, $n = 1, \ldots, 10000$, we would expect convergence to the following as $n \to \infty$ →

$$\Rightarrow \frac{1}{n}\left[h(U_1) + \ldots + h(U_n)\right] \approx E[U_1] = \int_{-\infty}^{\infty} x\,dx = \int_0^1 x\,dx = \left[\frac{x^2}{2}\right]_0^1 = \frac{1}{2}$$

↳ U_n iid for $h(x) = x$

∴ for $n \to \infty$, we would expect $\frac{1}{n}\sum_{K=1}^{n} U_K$ to converge to $\frac{1}{2}$

↳ for $\frac{1}{n}\sum_{K=1}^{n} U_K^5$, $n = 1, \ldots, 10000$, we would expect convergence to the following as $n \to \infty$ →

$$\Rightarrow \frac{1}{n}\left[h(U_1) + \ldots + h(U_n)\right] \approx E[U_1^5] = \int_{-\infty}^{\infty} x^5\,dx = \int_0^1 x^5\,dx = \left[\frac{x^6}{6}\right]_0^1 = \frac{1}{6}$$

↳ U_n iid for $h(x) = x^5$

∴ for $n \to \infty$, we would expect $\frac{1}{n}\sum_{K=1}^{n} U_K^5$ to converge to $\frac{1}{6}$

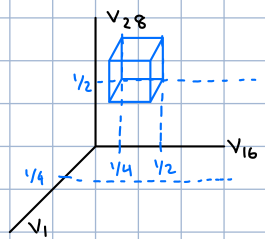
∴ we also see that LLN holds for our PRNGs as the converge to $\frac{1}{2}$ and $\frac{1}{6}$ respectively, except for PRNG A which is not large enough as we chose $m = 173$.

3. If we group pseudo-random sequence $U_1, U_2, \ldots, U_{nd}$ into disjoint blocks of length d, say $(U_1, \ldots, U_d)$, $(U_{d+1}, \ldots, U_{2d}), (U_{(n-1)d}, \ldots, U_{nd})$, then these d-dimensional blocks or random vectors should be uniformly distributed over the d-dimensional unit-cube $(0,1)^d$. For this problem, consider the following subset of the 30-dimensional unit cube:

$$E = \{(U_1, \ldots, U_{30}) \in (0,1)^{30} : U_1 \in (0, 1/4),\ U_{16} \in (1/4, 1/2),\ U_{28} \in (1/2, 1)\}$$

(a) Suppose $V_1, \ldots, V_{30}$ are iid uniform$(0,1)$. What is $P((V_1, \ldots, V_{30}) \in E)$? Notice that the event $(V_1, \ldots, V_{30}) \in E$ is just the event that $V_1 < 1/4$, $1/4 < V_{16} < 1/2$, and $1/2 < V_{28}$ simultaneously.

$\Rightarrow P((V_1, \ldots, V_3) \in E) = P(V_1 < 1/4) \cap P(1/4 < V_{16} < 1/2) \cap P(V_{28} > 1/2)$



$\Rightarrow$ pdf of uniform distribution $\rightarrow f(v) = 1$

$\Rightarrow P(V_1 < 1/4) = \int_{-\infty}^{1/4} f(v)\,dv = \int_0^{1/4} 1\,dv = \frac{1}{4}$

$\Rightarrow P(1/4 < V_{16} < 1/2) = \int_{1/4}^{1/2} f(v)\,dv = \int_{1/4}^{1/2} 1 = \frac{1}{4}$

$\Rightarrow P(V_{28} > 1/2) = \int_{1/2}^{\infty} f(v)\,dv = \int_{1/2}^{1} 1 = \frac{1}{2}$

$\Rightarrow P(V_1 < 1/4) \cap P(1/4 < V_{16} < 1/2) \cap P(V_{28} > 1/2) = \left(\frac{1}{4}\right)\left(\frac{1}{4}\right)\left(\frac{1}{2}\right) = \frac{1}{32}$

$\therefore P((V_1, \ldots, V_3) \in E) = \frac{1}{32}$

(b) For each of the PRNGs (A-D) defined above, generate a sequence of 10000 30-dimensional uniform pseudo-random vectors. What fraction of your 10,000 pseudo-random vectors are in set E?

↳ from attached MATLAB code, we can generate the fractions of the vectors in set E for each pseudo-random number generator (A-D) →

$\therefore$ fraction of PRNG A $\in$ E = 0.0349
fraction of PRNG B $\in$ E = 0.0324
fraction of PRNG C $\in$ E = 0
fraction of PRNG D $\in$ E = 0.0332

(c) For each PRNG in part (b), let T be the number of vectors (out of 10000) that are in the set E. If the PRNG was truly iid uniform, what would be the distribution of T? Use this knowledge to construct a test statistic to test the null hypothesis that 'the PRNG generates iid uniform (0,1) random variables'. Report your p-value.

↳ if the PRNG was truly uniform iid, $T \sim \text{Binomial}(10000, 1/32)$ → Test Statistic

↳ Hypothesis Testing: PRNG A

$\Rightarrow$ p-value $= P(T \geq T_{obs})$

$T_{obs} = 0.0349 \times 10000 = 349$ → use binocdf

p-value $= P(T \geq 349) = 2[1 - P(T < 349)]$
$= 2[0.018]$
$= 0.036$ [p-value < 0.05, must reject null hypothesis]

↳ Hypothesis Testing: PRNG B

$\Rightarrow T_{obs} = 0.0324 \times 10000 = 324$

p-value $= P(T \geq 324) = 2[1 - P(T < 324)]$
$= 2[0.2437]$
$= 0.4874$ [p-value high, can accept null hypothesis]

↳ Hypothesis Testing: PRNG C

$\Rightarrow T_{obs} = 0$

p-value $= P(T \leq 0) = 2[1.3096 \times 10^{-138}]$
$= 2.619 \times 10^{-138}$ [p-value < 0.05, must reject null hypothesis]

↳ Hypothesis Testing: PRNG D

$\Rightarrow T_{obs} = 0.0332 \times 10000 = 332$

p-value $= P(T \geq 332) = 2[1 - P(T < 332)]$
$= 2[0.1257]$
$= 0.2514$ [p-value high, can accept null hypothesis]

HOMEWORK 2

Anvita Bhagavathula

1. Consider the pmf $p(x) = c\sqrt{x}\,\mathbb{1}_{x \in \{1, \ldots, 50\}}$

(a) Calculate the numerical value of $c \rightarrow$

↳ the sum of all values of a pmf must equal $1 \rightarrow$

$$\Rightarrow \sum_{x=1}^{50} p(x) = \sum_{x=1}^{50} c\sqrt{x} = 1$$

$$\Rightarrow c\sqrt{1} + c\sqrt{2} + c\sqrt{3} + \ldots c\sqrt{50} = 1$$

$$\Rightarrow c\left[\sqrt{1} + \sqrt{2} + \sqrt{3} + \ldots + \sqrt{50}\right] = 1$$

$$\Rightarrow c \sum_{x=1}^{50} \sqrt{x} = 1$$

$$\Rightarrow c = \frac{1}{\sum_{x=1}^{50} \sqrt{x}} = \frac{1}{\sqrt{1} + \ldots + \sqrt{50}}$$

↳ the sum $\sqrt{1} + \ldots + \sqrt{50}$ can be found using MATLAB →

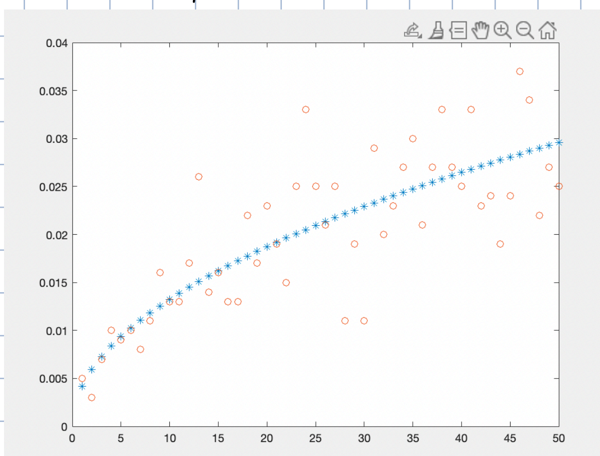
$$\Rightarrow \sum_{x=1}^{50} \sqrt{x} = 239.03$$

$$\Rightarrow c = \frac{1}{239.03} = 0.0042$$

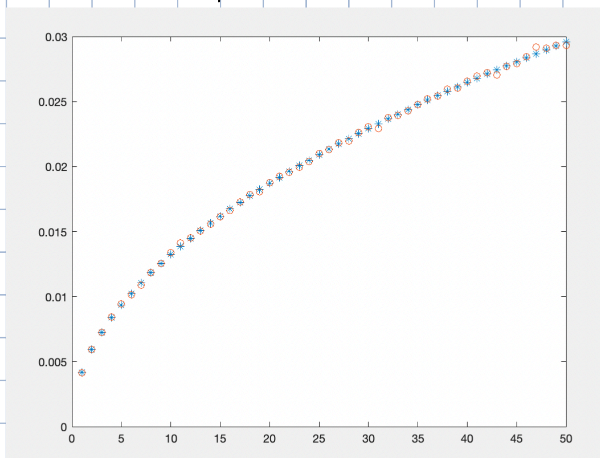
$\therefore$ $c = 0.0042$

(b) For $n = 10^3$, generate $X_1, \ldots, X_n$ iid with common pmf p. Let $\hat{p}(x) = \#\{k : X_k = x\}/n$ be the empirical pmf of $X_1, \ldots, X_n$. Make a graph comparing $p(x)$ and $\hat{p}(x)$. Repeat with $n = 10^6$.

Graph comparing $p(x)$ and $\hat{p}(x)$ for $n = 10^3$ samples



Graph comparing $p(x)$ and $\hat{p}(x)$ for $n = 10^6$ samples



$\therefore$ We observe that for a larger sample size of $n = 10^6$, $\hat{p}(x)$ more closely / accurately estimates $p(x)$.

2. Consider the pdf $f(x) = \frac{1}{2\sqrt{x}} \mathbb{1}\{0 < x < 1\}$

(a) How would you convert a uniform (0,1) rv U into a rv X that had pdf f?

↳ $U \sim \text{unif}[0,1]$

↳ $X \sim f(x)$

↳ we can apply the THM we covered in class which states the following →

↳ Suppose F is a CDF. Suppose U is unif [0,1]. Then an r.v. X given by $X = F^{-1}(U)$ has a cdf F.

↳ if X has pdf $f(x)$, it has cdf $F(x)$ given by →

$$\Rightarrow F(x) = \int_0^x f(x)\,dx = \int_0^x \frac{1}{2\sqrt{x}}\,dx = \sqrt{x}$$

↳ Since $X = F^{-1}(U)$, we know $U = F(X)$. Hence we can convert $U \sim \text{unif}[0,1]$ into an r.v. X through the following expression →

$\Rightarrow U = F(X) = \sqrt{X}$

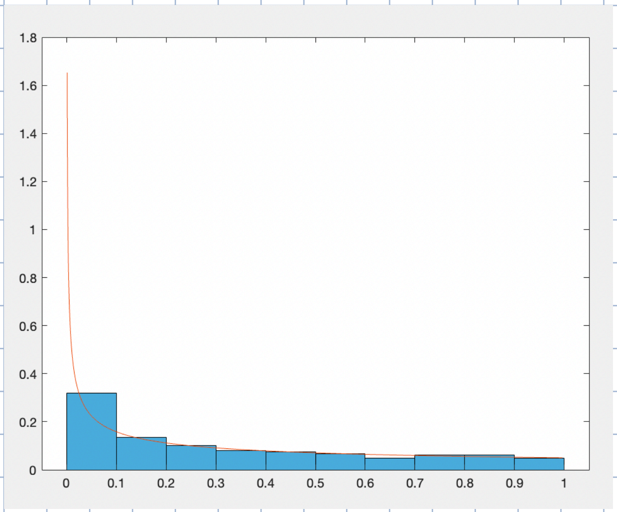
$\Rightarrow U = \sqrt{X}$

$\Rightarrow X = U^2$

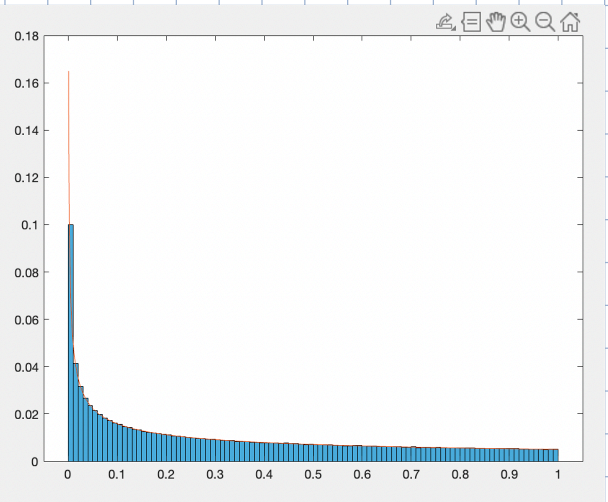
$\therefore$ we can convert via $X = F^{-1}(U) = U^2$

(b) Generate a sample of size 10^3 from f and compare a plot of f to a histogram of your samples. Repeat with a sample of size 10^6.

Graph of plot of f and histogram of samples of 10^3 samples.



Graph of plot of f and histogram of samples of 10^6 samples.



3. Consider the pdf $f(x) = \frac{1}{c}(\sin(10x))^2 |x^3 + 2x - 3| \mathbb{1}\{x \in (-1,0) \cup (1/2, 1)\}$

(a) Use rejection sampling to generate a sample size of 10^6, and approximate the numerical value of c →

↳ we know that $g(x) = c f(x) = (\sin(10x))^2 |x^3 + 2x - 3|$

↳ to run a rejection sampling scheme we first need to decide on a region E that covers $g(x) = c f(x)$. [rectangle]

↳ to do so, let us consider the max values of $g(x)$ →

↳ we know $x \in (-1,0) \cup (1/2, 1)$, hence we will generate samples for $x \in (-1,1)$

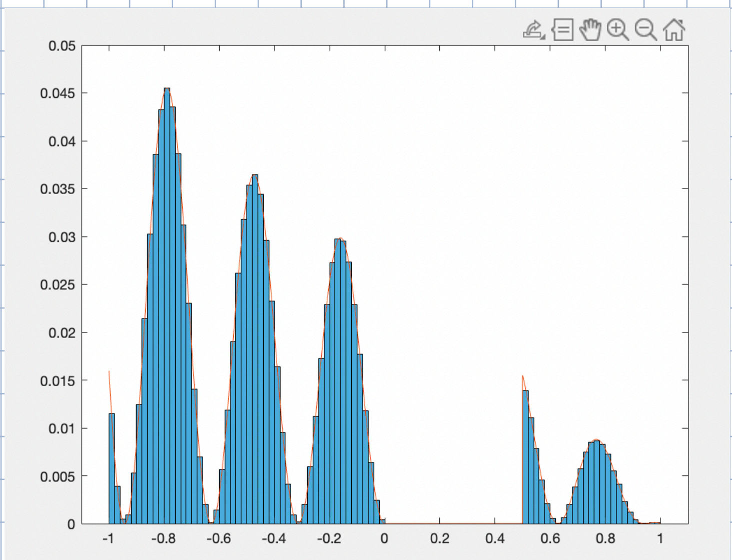
↳ the max y values can be found through $f(x)$. The max of $\sin(10x)^2$ is always 1. The max of $|x^3 + 2x - 3|$ exist at $x = -1$, $x = 1$ given by $y = 6$, $y = 0$ respectively.

↳ hence the rectangle region E we will use is $[-1, 1] \times [0, 6]$ while also considering the region $x \in (0, 1/2)$ where $g(x)$ is 0.

↳ using MATLAB (see code attached), we can calculate the acceptance ratio of the samples we run the rejection sampling scheme on to calculate c.

∴ numerical value of $c = 2.2247$

(b) Compare a plot of f to a histogram of your samples →



4. Consider the pdf $f(x) = \frac{1}{C}\frac{1}{2\sqrt{x}}\sqrt{\left|\sin\left(\frac{10}{1+|\log(x)|}\right)\right|}\,\mathbb{1}_{\{0<x<1\}}$

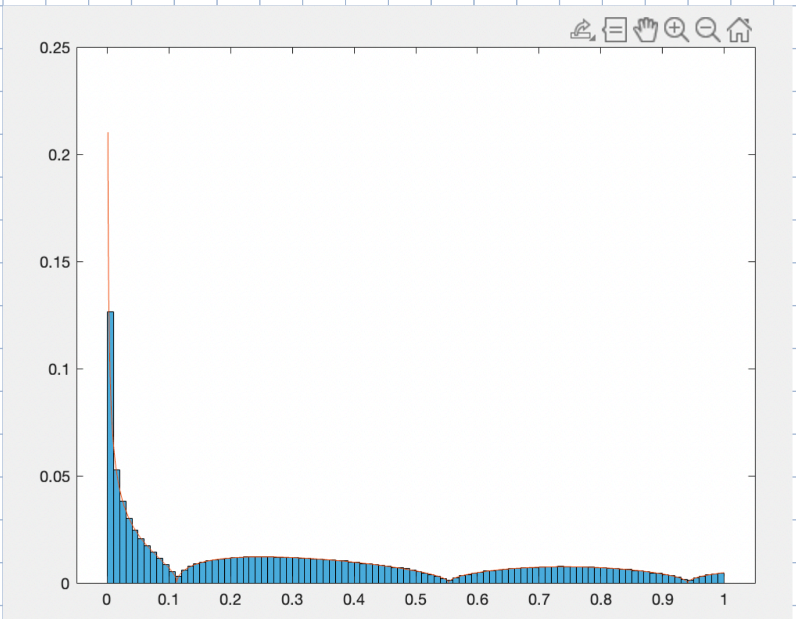
(a) Use rejection sampling to generate a sample of size 10^6 from f, and approximate the numerical value of C.

↳ to create a region E to run rejection sampling on our samples, we need to generate y values between 0 and $f(x) = \frac{1}{2\sqrt{x}}$ (since the sine term will have a max value of 1) while $x \in (0,1)$.

↳ we can employ the inverse CDF method to generate uniform random samples for x, y in MATLAB (see code attached) and calculate the acceptance ratio of the samples.

∴ numerical value of C = 0.7645.

(b) Compare a plot of f to a histogram of your samples →



HOMEWORK 4

Anvita Bhagavathula

1. Define the function $h(\vec{x}) = \left|\sin\left(2\pi x_1 \sum_{i=1}^{100} x_i\right)\right| \left(\cos\left(2\pi x_2 \sum_{j=1}^{100} x_j\right)\right)^2$ for $\vec{x} = (x_1, \dots, x_{100}) \in \mathbb{R}^{100}$. For each of the following problems, $n = 10^5$ samples, and simply take $f(\vec{x})$ as the uniform distribution on appropriare regions.

(a) Compute an approximate 95% confidence interval for the value of $\int_{[0,1]^{100}} h(\vec{x})\, d\vec{x}$

↳ see attached MATLAB code

$\therefore$ CI $= [0.3184, 0.3221]$

(b) Compute an approximate 95% confidence interval for the value of $\int_{[0,1.05]^{100}} h(\vec{x})\, d\vec{x}$

↳ see attached MATLAB code

↳ for this condition, we consider $f(x) = \dfrac{1}{1.05^{100}}$

↳ $v = \displaystyle\int \frac{h(\vec{x})}{f(\vec{x})} f(\vec{x})\, d\vec{x}$

$\therefore$ CI $= [41.5154, 42.0029]$

(c) Compute an approximate 95% confidence interval for the value of $\int_{[0,1]^{100} \cap B} h(\vec{x})\, d\vec{x}$ where B is the ball of radius 6 in $\mathbb{R}^{100}$.

↳ we know that the following is true →

$\Rightarrow \displaystyle\int_{[0,1]^{100} \cap B} h(\vec{x})\, d\vec{x} = \int_{[0,1]^{100}} h(\vec{x})\, \mathbb{1}_{\{\vec{x} \in B\}}\, d\vec{x} = \int_{[0,1]^{100}} h(\vec{x})\, \mathbb{1}_{\{\|x\| \leq 6\}}\, d\vec{x}$

↳ See attached MATLAB code

$\therefore$ CI $= [0.2578, 0.2615]$

(d) What goes wrong if you try to do the previous problem using a ball of radius 4?

↳ if we attempt the above problem with a ball of radius 4, we see that our mean $= 0$, standard deviation $= 0$, and CI $= [0, 0]$

↳ hence we can conclude that the minimum length for our uniformly generated samples $\vec{x}_i$ on $[0,1]^{100}$ is greater than 4.

2. In Problem 1c-d you had to integrate a function over $[0,1]^{100} \cap B_r$ where B_r is the ball of radius r in $\mathbb{R}^{100}$. Using uniform samples for your Monte Carlo, you have found that this worked pretty easily for $r=6$ and not so easily for $r=4$.

(a) Use the central limit theorem to approximate the volume of $[0,1]^{100} \cap B_r$ for $r>0 \rightarrow$

↳ Volume of $[0,1]^{100} \cap B_r$ for $r>0 \rightarrow$

$$\Rightarrow V = \int_{[0,1]^{100} \cap B_r} d\vec{x} = \int_{[0,1]^{100}} \mathbb{1}_{\{|\vec{x}| \le r\}} d\vec{x} = E[\mathbb{1}_{\{|\vec{x}| \le r\}}]$$

↳ if we consider $\vec{x} = (x_1, \dots, x_{100})$ where $x_i \sim \text{unif}[0,1]$ iid.

$$\Rightarrow E[\mathbb{1}_{\{|\vec{x}| \le r\}}] = P(x_1^2 + \dots + x_{100}^2 \le r^2)$$

↳ we can use CLT to approximate the iid random variables →

$$\Rightarrow (x_1^2 + \dots + x_{100}^2) = \sum_{k=1}^{n} x_k^2 \approx N(n\mu, n\sigma^2) \text{ for } n \text{ large}$$

$$\Rightarrow P(x_1^2 + \dots + x_{100}^2 \le r^2) = P(N(n\mu, n\sigma^2) \le r^2)$$

↳ Now let us determine μ, σ^2 for $(X_1^2, \dots, X_{100}^2) \rightarrow$

$$\Rightarrow \mu = E[X_i^2] = \int_{-\infty}^{\infty} x^2 \, \mathbb{1}_{\{x \in [0,1]\}} dx = \int_0^1 x^2 dx = \left[\frac{x^3}{3}\right]_0^1 = \frac{1}{3} \quad \text{(by LLN)}$$

$$\Rightarrow \sigma^2 = E[X_i^4] - (E[X_i^2])^2 = \int_{-\infty}^{\infty} x^4 \, \mathbb{1}_{\{x \in [0,1]\}} dx - \left(\frac{1}{3}\right)^2 = \frac{1}{5} - \frac{1}{9} = \frac{4}{45}$$

↳ Hence we obtain →

$$\Rightarrow P(N(n\mu, n\sigma^2) \le r^2) = P\left(N\left(100\left(\tfrac{1}{3}\right), 100\left(\tfrac{4}{45}\right)\right) \le r^2\right)$$

$$\Rightarrow P\left(\frac{N(100/3, 400/45) - 100/3}{\sqrt{100}\sqrt{4/45}} \le \frac{r^2 - 100/3}{\sqrt{100}\sqrt{4/45}}\right) = P\left(Z \le \frac{r^2 - 100/3}{\sqrt{100}\sqrt{4/45}}\right)$$

∴ The volume is approximately $\Phi\left(\dfrac{r^2 - 100/3}{10\sqrt{4/45}}\right)$

(b) If we try to solve 1c-d by using Monte Carlo samples from $U(0,1)^{100}$, approximately what fraction of our samples will land in the region of integration for Problem 1c? And approximately what fraction for Problem 1d?

↳ for $r=6$ →

$$\Phi\left(\frac{36 - 100/3}{10\sqrt{4/45}}\right) = 0.8145 = 81.45\% \text{ of samples will land in region}$$

↳ for $r=4$ →

$$\Phi\left(\frac{16 - 100/3}{10\sqrt{4/45}}\right) = 3.0539 \times 10^{-9} = 3.0539 \times 10^{-7}\% \text{ of samples will land in region}$$

(c) Estimate the fractions in part (b) by Monte Carlo using 10^6 samples for the case $r=6$. Report your 95% confidence intervals. How does Monte Carlo approximation compare to your CLT approximations?

↳ See attached MATLAB code

↳ the CLT approximation falls within the range of the Monte Carlo approximations

∴ estimated fraction = 0.8154 = 81.54%

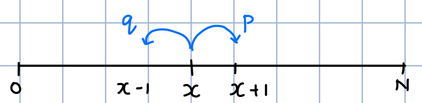
95% confidence interval = [0.7976, 0.8332]

HOMEWORK 5

Anvita Bhagavathula

1. Consider the following game →
   - ↳ you pay \$1 to play
   - ↳ I roll a six-sided die, you roll a six-sided die
   - ↳ If your die beats mine, or both = 6 or 1, I give you \$2. Otherwise I keep your \$1.
   - ↳ You begin the game with \$20 and plan to stop at \$25 or if you run out of money
   - ↳ What is the probability of reaching \$25? What are your expected winnings? Would you play this game?

↳ Gambler's Ruin Setup →



$\Rightarrow P_i = pP_{i+1} + qP_{i-1}$

$\Rightarrow p$ = probability of second die rolling higher number than first die or both dies rolling 6 or 1.

$\Rightarrow q = 1 - p$

↳ first let us determine p and q →

$\Rightarrow$ total # of outcomes rolling 2 dice $= 6 \times 6 = 36$

$\Rightarrow$ # of outcomes when my die higher than yours or both of the dice are 6 or 1 =

(1, 1) (1, 2) (1, 3) (1, 4) (1, 5) (1, 6)
(2, 3) (2, 4) (2, 5) (2, 6)
(3, 4) (3, 5) (3, 6)
(4, 5) (4, 6)
(5, 6)
(6, 6)

$\Rightarrow p = \frac{17}{36}$, $q = 1 - \frac{17}{36} = \frac{19}{36}$

↳ P(reaching \$25 before \$0 | starting at \$20)

$\Rightarrow P_{25} = 1$

$P_0 = 0$

$$\Rightarrow \begin{bmatrix} P_0 = 0 \\ -qP_0 + P_1 - pP_2 = 0 \\ \vdots \\ P_{25} = 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -q & 1 & -p & \cdots & 0 \\ 0 & -q & 1 & -p \;\cdots & 0 \\ \vdots & & & & \vdots \\ 0 & \cdots & \cdots & & 1 \end{bmatrix} \begin{bmatrix} P_0 \\ P_1 \\ \vdots \\ P_{25} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

$\Rightarrow$ P(reaching \$25 before \$0 | starting at \$20) for $S_n = S_0 + \sum_{i=1}^{n} X_i$ $= P(S_T = 25 \mid S_0 = 20) = P_{20}$

$\Rightarrow$ P(reaching \$25 before \$0 | starting at \$20) $= P_{20} = 0.5147$ (see attached MATLAB code)

$\therefore P_{20} = P(21)$ (in MATLAB) $= 0.5452$

↳ Expected winnings for this game →

$\Rightarrow E[x] =$ P(running out of money) $\cdot (-20)$ + P(reaching \$25)(5)

$\Rightarrow$ P(running out of money) = 1 − P(reaching \$25 before \$0)

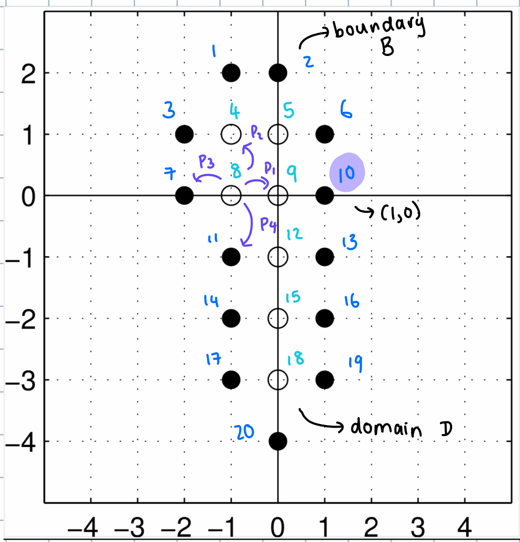
$= 1 - 0.5452$

$= 0.4548$

$\therefore E[x] = (0.4548)(-20) + (0.5452)(5) = -\$6.37$

$\therefore$ Since the expected winnings are less than \$0, I wouldn't play this game

2. The figure below shows a region in the 2d integer lattice. Let D denote the interior of the region (open circles) and let B denote the boundary (filled circles). Consider a symmetric simple random walk $S = (S_0, S_1, \ldots)$ starting at an interior point. Define →
$u(x,y) = P(S \text{ first exits } D \text{ at } (1,0) \mid S_0 = (x,y))$



(a) Use MATLAB to solve the system of equations to compute $u(x,y)$ for every $(x,y) \in D$.

⇒ let $x_i := P(S_n \text{ reaches } (1,0) \text{ before other } D \mid S_0 = (x,y))$

⇒ symmetric random walk means $p_1 = p_2 = p_3 = p_4 = 1/4$

⇒ $$\begin{bmatrix} 1 & 0 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ -P_2 & 0 & -P_3 & 1 & -P_4 & \cdots \\ \vdots & & & & & \vdots \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ \vdots \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \vdots \\ 1 \\ 0 \end{bmatrix}$$ ← at 10th row

↳ see attached MATLAB code for full vector $\vec{x}$

∴ $u(x,y)$ for every $(x,y) \in D$ is given by →

$u(-1,1) = 0.0452$ $u(0,-1) = 0.0847$

$u(0,1) = 0.0904$ $u(0,-2) = 0.0226$

$u(-1,0) = 0.0904$ $u(0,-3) = 0.0056$

$u(0,0) = 0.3164$

(b) Test your answer for $u(0,0)$ by Monte Carlo integration. Repeatedly start a random walk at $(0,0)$ and run until it first exits D. Compute the fraction of the time the walk exits at $(1,0)$. Give an approximate 95% confidence interval for $u(0,0)$, with sample size $n = 10^5$.

↳ see attached MATLAB code

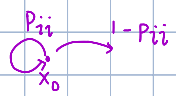
∴ fraction of the time the walk exits at $(1,0) = 0.3164$

∴ 95% confidence interval for $u(0,0) = [0.3134, 0.3193]$

HOMEWORK 6

Anvita Bhagavathula

1. Suppose that X is a Markov Chain starting at state $X_0 = i$. Assume that $p = P_{ii} > 0$. Let T be the first time the Markov chain leaves the initial state i, that is, $T = \inf\{n \geq 1 : X_n \neq i\}$. What is the distribution of T?



probability that Markov chain ends up at $X_n = i$ given $X_0 = i \rightarrow P(X_n = i | X_0 = i) = P_{ii}$
probability that Markov chain ends up at $X_n \neq i$ given $X_0 = i \rightarrow P(X_n \neq i | X_0 = i) = 1 - P_{ii}$

$\Rightarrow$ $T = \inf\{n \geq 1 : X_n \neq i\}$ $\rightarrow$ we want to find $P(T = n)$ which is the probability that the chain leaves the state X_0 at the n-th time-step. This means that until the n-th time step, the Markov chain remains at state $i \rightarrow$

$\Rightarrow$ for $n = 2$:
$X_0 = i,\ X_1 = i,\ X_2 \neq i$
$P(T = 2) = (P_{ii})(1 - P_{ii})$

$\Rightarrow$ for $n = 3$:
$X_0 = i,\ X_1 = i,\ X_2 = i,\ X_3 \neq i$
$P(T = 3) = (P_{ii})^2(1 - P_{ii})$

$\vdots$

$\Rightarrow$ for $n = n$
$X_0 = i,\ X_1 = i, \ldots, X_{n-1} = i,\ X_n \neq i$
$P(T = n) = (P_{ii})^{n-1}(1 - P_{ii})$

$\therefore$ T is a geometric distribution where $P(T = n) = (P_{ii})^{n-1}(1 - P_{ii})$

2. Toss a fair coin repeatedly until two successive heads appear. Find the expected number of tosses needed. Let X_n be the cumulative number of successive heads at the n-th toss. We can assume X_n takes value in state space $\{0,1,2\}$ and $\{2\}$ is the absorbing state. Then use first-step analysis to solve the problem.

$\Rightarrow$ let $(Y_n)_{n \geq 1}$ be i.i.d counter variables with $Y_n = \begin{cases} 1 & \text{if n-th flip is heads} \\ 0 & \text{if n-th flip is tails} \end{cases}$

$\Rightarrow$ $P(Y_n = 1) = P(Y_n = 0) = 1/2$

$\Rightarrow$ we can set up a transition matrix for the Markov chain X_n using the probabilities defined by the i.i.d counter variables above. For example $\rightarrow$

$$\left.\begin{aligned}
P_{00} &= P(X_{n+1} = 0 | X_n = 0) = P(Y_n = 0) = 1/2 \\
P_{01} &= P(X_{n+1} = 1 | X_n = 0) = P(Y_n = 1) = 1/2 \\
&\vdots \\
P_{12} &= P(X_{n+1} = 2 | X_n = 1) = P(Y_n = 1) = 1/2 \\
&\vdots \\
P_{21} &= P(X_{n+1} = 1 | X_n = 2) = 0 \\
P_{22} &= P(X_{n+1} = 2 | X_n = 2) = 1 \quad \text{(Two consecutive H)}
\end{aligned}\right\}$$

$$P = \begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 0 & 1 \end{bmatrix} \end{array} \begin{array}{c} 0 \\ 1 \\ 2 \end{array} \quad \left[\begin{array}{c} K_0 \\ K_1 \\ K_2 \end{array} \right.$$

$\Rightarrow$ let us define $T = \min\{n \geq 0 | X_n = 2\}$. We want to find $E[T]$ (apply first step analysis).

$\Rightarrow$ define K_i^2 = expected number of states to hit $X_n = \{2\}$ given initial state i

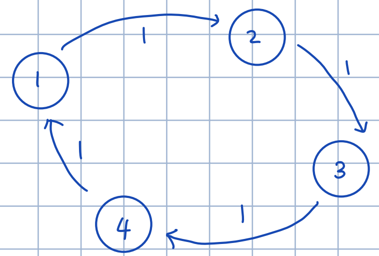
$$\left.\begin{aligned}
K_0^2 &= 1 + 1/2 K_0^2 + 1/2 K_1^2 \\
K_1^2 &= 1 + 1/2 K_0^2 + 1/2 K_2^2 \\
K_2^2 &= 0 \quad \text{(we are already at } \{2\})
\end{aligned}\right\} \Rightarrow \begin{aligned} 1/2 K_0^2 &= 1 + 1/2 K_1^2 \\ K_1^2 &= 1 + 1/2 K_0^2 \end{aligned}$$

$\Rightarrow$ $1/2 K_0^2 = 1 + 1/2(1 + 1/2 K_0^2) \Rightarrow 1/4 K_0^2 = 1 + 1/2 \Rightarrow K_0^2 = 6$

$\therefore$ $E[T] = K_0^2$ since we start with 0 consecutive heads and want to end up at 2 consecutive heads. The expected number of flips is $E[T] = 6$.

3. Even though in nearly all of the meaningful models of Markov Chains irreducibility and aperiodicity conditions hold, it is worthwhile to understand what happens when they fail

(a) When the aperiodicity assumption fails, an irreducible Markov Chain may not converge to the stationary distribution. Please construct a very simple example of a finite state, periodic Markov Chain to illustrate this.



=> transition matrix for Markov chain X_n on left →

=> $P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$

=> to verify that the assumption fails, first let us compute the stationary distribution of X_n.

=> stationary distribution $\pi(x) = [\pi(1), \pi(2), \pi(3), \pi(4)]$

=> $\pi(1) + \pi(2) + \pi(3) + \pi(4) = 1$

=> $P\pi(x) = \pi(x)$ → $\left.\begin{aligned} P_{41}\,\pi(4) &= \pi(1) = \pi(4) \\ P_{12}\,\pi(1) &= \pi(2) = \pi(1) \\ P_{23}\,\pi(2) &= \pi(3) = \pi(2) \\ P_{34}\,\pi(3) &= \pi(4) = \pi(3) \end{aligned}\right\}$ $\begin{aligned} &\pi(1) + \pi(2) + \pi(3) + \pi(4) = 1 \\ &\pi(1) = \pi(2) = \pi(3) = \pi(4) \end{aligned}$

=> solving the above, we obtain $\pi(1) = \pi(2) = \pi(3) = \pi(4) = 1/4$

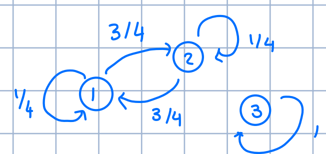
=> now let us examine what X_n converges to →

=> $\left.\begin{aligned} P(X_1 = 1 \mid X_0 = 1) &= 0 \\ P(X_2 = 1 \mid X_0 = 1) &= 0 \\ P(X_3 = 1 \mid X_0 = 1) &= 0 \\ P(X_4 = 1 \mid X_0 = 1) &= 1 \\ &\vdots \\ P(X_8 = 1 \mid X_0 = 1) &= 1 \\ &\vdots \end{aligned}\right\}$

∴ Because $X_n = 1$ for $n = 4, 8, 16, \ldots$, it never converges to $\pi(1)$ as $n \to \infty$. Hence, X_n does not converge to a stationary distribution when the aperiodity assumption fails.

(b) When the irreducibility assumption fails, a Markov Chain may have infinitely many stationary distributions. Consider the Markov chain with the transition matrix below. For this Markov chain find all the stationary distributions →

=> $P = \begin{bmatrix} 1/4 & 3/4 & 0 \\ 3/4 & 1/4 & 0 \\ 0 & 0 & 1 \end{bmatrix}$



=> stationary distribution $\pi(x) = [\pi(1), \pi(2), \pi(3)]$

=> $\pi(x)P = \pi(x)$

=> $\pi(1) + \pi(2) + \pi(3) = 1$

=> $\left.\begin{aligned} 1/4\,\pi(1) + 3/4\,\pi(2) &= \pi(1) \\ 3/4\,\pi(1) + 1/4\,\pi(2) &= \pi(2) \\ \pi(3) &= \pi(3) \end{aligned}\right\}$ $\begin{aligned} &3/4\,\pi(1) = 3/4\,\pi(2) \\ &\Rightarrow \pi(1) = \pi(2) \end{aligned}$

=> $2\pi(1) + \pi(3) = 1$

$\pi(3) = 1 - 2\pi(1)$

∴ the stationary distributions are given by $\pi(x) = [\pi(1), \pi(1), 1 - 2\pi(1)]$ where $1 - 2\pi(1) \geq 0 \iff \pi(1) \leq 1/2$

4. Consider an irreducible and aperiodic Markov Chain with state space $\Omega = \{1,2,3\}$ and transition probability matrix given below.

(a) Compute the stationary distribution of this Markov Chain →

$\Rightarrow$ $P = \begin{bmatrix} 0 & 1/3 & 2/3 \\ 1/3 & 0 & 2/3 \\ 1 & 0 & 0 \end{bmatrix}$

$\Rightarrow$ Stationary distribution $\pi = [\pi(1), \pi(2), \pi(3)]$

$\Rightarrow$ $\pi P = \pi$

$\Rightarrow$ $\pi(1) + \pi(2) + \pi(3) = 1$

$$\left.\begin{aligned} 1/3\,\pi(2) + \pi(3) &= \pi(1) \\ 1/3\,\pi(1) &= \pi(2) \\ 2/3\,\pi(1) + 2/3\,\pi(2) &= \pi(3) \end{aligned}\right\} \quad \begin{aligned} &\pi(1) = 3\pi(2) \\ &2\pi(2) + 2/3\,\pi(2) = \pi(3) \\ &8/3\,\pi(2) = \pi(3) \end{aligned}$$

$\Rightarrow$ $3\pi(2) + \pi(2) + 8/3\,\pi(2) = 1$

$20/3\,\pi(2) = 1$

$\pi(2) = 3/20$, $\pi(1) = 9/20$, $\pi(3) = 8/20$

$\therefore$ $\pi = [9/20, 3/20, 8/20]$

(b) What is the long run proportion time that the Markov chain spends on state 1?

$\Rightarrow$ $\frac{1}{n}\sum_{k=1}^{n} \mathbb{1}\{x_k = 1\}$ as $n \to \infty$ $= \pi(1)$ [THM 3.5]

$\therefore$ the proportion of time is $9/20$

(c) Determine the limiting matrix $\lim_{n\to\infty} P^n$ and explain your answer.

$\Rightarrow$ components of the matrix P^n →

$\Rightarrow$ $[P^n]_{xy} = P(X_n = y \mid X_0 = x)$

$\Rightarrow$ $\lim_{n\to\infty} [P^n]_{xy} = \lim_{n\to\infty} P(X_n = y \mid X_0 = x) = \pi(y)$ as per THM 3.5 which states that regardless of the initial distribution X_0, for $x \in \Omega$ we have $P(X_n = x) \longrightarrow \pi(x)$ as $n \to \infty$.

$\therefore$ the limiting matrix $\lim_{n\to\infty} P^n$ is equal to $\begin{bmatrix} \pi(1) & \pi(2) & \cdots & \pi(n) \\ \pi(1) & \pi(2) & \cdots & \pi(n) \\ \vdots & & & \\ \pi(1) & \pi(2) & \cdots & \pi(n) \end{bmatrix}$

(d) Lucky Harry bets on the Markov Chain. At every step

(i) if $X_n = 1$, Lucky Harry wins \$1

(ii) if $X_n = 2$, Lucky Harry wins \$2

(iii) if $X_n = 3$, Lucky Harry loses \$1

What is Harry's average net profit?

$\Rightarrow$ in the long run, the $E[\text{Profit}] = \sum \text{profit} \cdot P(\text{profit in long run})$

$\Rightarrow$ P(profit in long run is given by a stationary distribution

$\Rightarrow$ $E[\text{Profit}] = \pi(1)[1] + \pi(2)[2] + \pi(3)[-1]$

$= \frac{9}{20} + \frac{6}{20} - \frac{8}{20} = \frac{7}{20}$

$\therefore$ Harry's net profit in the long run is \$ $7/20$

HOMEWORK 7

Anvita Bhagavathula

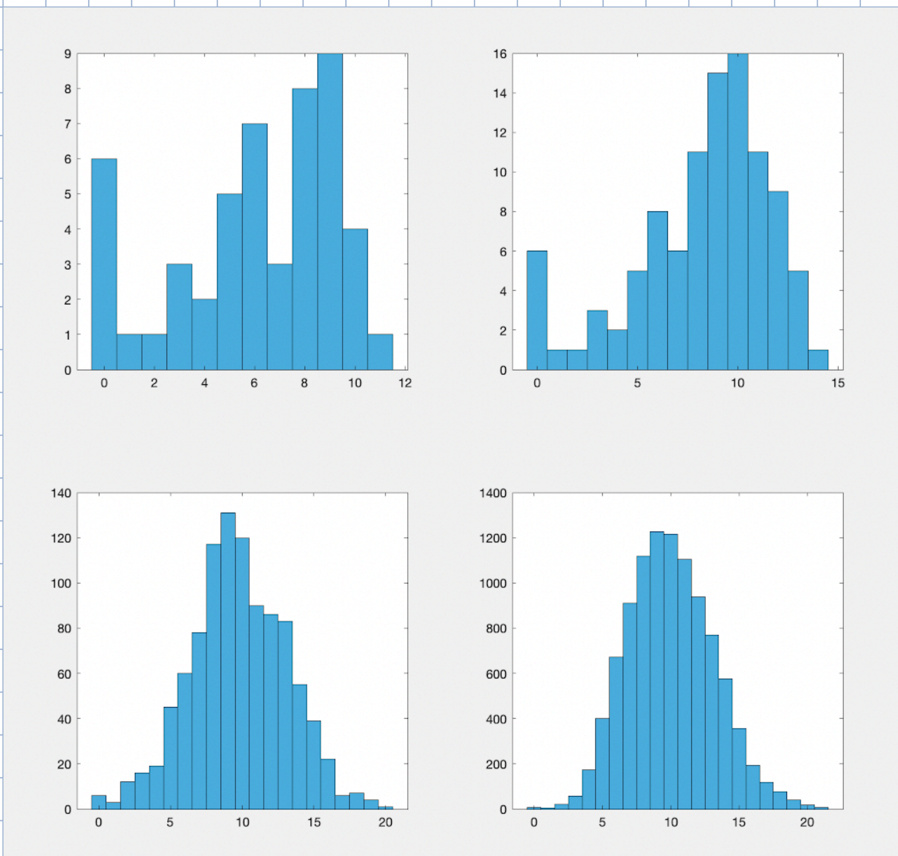
1. (Metropolis- Hastings) Suppose we wish to sample from the Poisson distribution with parameter 10. The target distribution is $\pi(x) = e^{-10}\frac{10^x}{x!}$ $x = 0, 1, 2, \ldots$ Suppose the proposal transition matrix Q is such that

$$Q_{xy} = \begin{cases} 0.5 & \text{if } y = x \pm 1 \\ 0 & \text{otherwise} \end{cases} \quad \text{for } x \geq 1$$

$$Q_{0y} = \begin{cases} 0.5 & \text{if } y = 1 \text{ or } 0 \\ 0 & \text{otherwise} \end{cases}$$

Build your metropolis - hastings algorithm for simulating π. Run your Markov Chain $\{X_0, \ldots, X_n\}$ starting from $X_0 = 0$, with $n = 10^4$.

(a) For $K = 50, 10^2, 10^3, 10^4$, draw a histogram of $\{X_1, \ldots, X_K\}$.



(b) Use your Markov Chain to estimate $P(X=2)$, $E[X]$ and $E[X^2]$ where X is the Poisson with parameter 10. You only need to report your estimates. Compare your estimates with the theoretical values. Do they look consistent to you?

↳ See attached MATLAB code

↳ Approximations from the Markov Chain →

$$\left.\begin{aligned} &\Rightarrow P(X = 10) = 0.1216 \\ &\Rightarrow E[X] = 9.9991 \\ &\Rightarrow E[X^2] = 109.901 \end{aligned}\right\} \text{approximations from the Markov Chain}$$

↳ Theoretical Values →

$\Rightarrow$ $P(X = 10) = e^{-10}\left(\frac{10^{10}}{10!}\right) = 0.1251$

$\Rightarrow$ $E[X] = \lambda = 10$

$\Rightarrow$ $E[X^2] = Var[X] + E[X]^2 = \lambda + \lambda^2 = 110$

$\therefore$ Estimates are pretty close to the theoretical values - they look consistent.

2. Ising Model

(a) For $R = C = 50$ and for each of the temperatures $T = 0.1, 1, 2, 3, 10$ run the Gibbs sampler for 1000 full sweeps. Initialise with iid 50% / 50% ± 1 entries as above. Take 'snap-shots' of the process after 250, 500, 750, 1000 sweeps. Make a figure showing all of the results.

↳ Conditional Probabilities → $X_i = \pm 1$

$$P(x_{ij} = 1 \mid \cdots) = \frac{P(x_{ij}=1, \ldots)}{P(\ldots)} = \frac{1}{Z_T} \exp\left(\frac{1}{T} \sum_{\{(i,j),(k,l)\}} x_{ij} x_{kl}\right) \frac{1}{P(\ldots)} = \frac{1}{Z_T} \exp\left(\frac{1}{T} \sum_{\{(i,j),(k,l)\}} x_{kl}\right) \frac{1}{P(\ldots)}$$

$$P(x_{ij} = -1 \mid \cdots) = \frac{P(x_{ij}=-1, \ldots)}{P(\ldots)} = \frac{1}{Z_T} \exp\left(\frac{1}{T} \sum_{\{(i,j),(k,l)\}} x_{ij} x_{kl}\right) \frac{1}{P(\ldots)} = \frac{1}{Z_T} \exp\left(\frac{-1}{T} \sum_{\{(i,j),(k,l)\}} x_{kl}\right) \frac{1}{P(\ldots)}$$

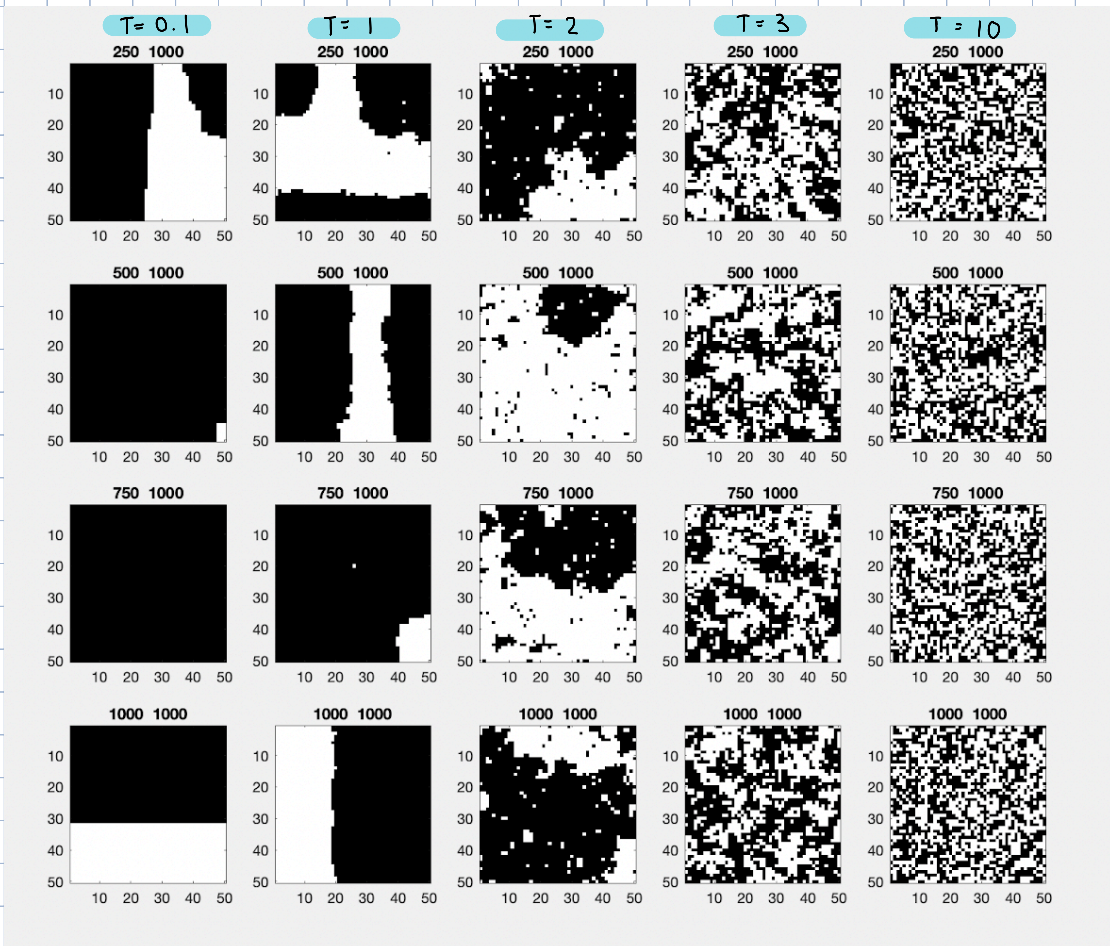
$$\Rightarrow P(x_{ij} = 1 \mid \cdots) + P(x_{ij} = -1 \mid \cdots) = 1$$

$$\Rightarrow \text{ratio} = \frac{P(x_{ij} = 1 \mid \cdots)}{P(x_{ij} = -1 \mid \cdots)} = \exp\left(\frac{2}{T} \sum_{\{(i,j),(k,l)\}} x_{kl}\right)$$

$$\left\{\begin{array}{l} \Rightarrow P(x_{ij} = 1 \mid \cdots) = \dfrac{\text{ratio}}{\text{ratio} + 1} = \dfrac{\exp\left(\frac{2}{T} \sum_{\{(k,l)\}} x_{kl}\right)}{\exp\left(\frac{2}{T} \sum_{\{(k,l)\}} x_{kl}\right) + 1} \\ \Rightarrow P(x_{ij} = -1 \mid \cdots) = \dfrac{1}{\text{ratio} + 1} = \dfrac{1}{\exp\left(\frac{2}{T} \sum_{\{(k,l)\}} x_{kl}\right) + 1} \end{array}\right\}$$

Conditional probabilities for the Gibbs sampler.

↳ See attached MATLAB code



∴ The Gibbs sampler seems to be working more accurately at higher temperatures.
Hence, the Ising Model converges at higher temperatures.

(b) Suppose you wanted to Monte Carlo approximate $\theta_T = E_T[\exp(X_{11} + X_{22})]$ for the case $R = C = 50$ at temperatures $T = 0.1, 1, 2, 3, 10$.

(i) Explain how you would use your code to do this and submit your estimates.

↳ create a vector that updates with the desired statistic $(\exp(x_{11} + x_{22}))$ at each sweep for given T

↳ take mean of the vector for every T.

↳ see attached MATLAB code

↳ estimates →

$\Rightarrow$ $T = 0.1 \rightarrow \theta_T = 7.3891$

$T = 1 \rightarrow \theta_T = 7.2796$

$T = 2 \rightarrow \theta_T = 2.0292$

$T = 3 \rightarrow \theta_T = 2.6684$

$T = 10 \rightarrow \theta_T = 2.3394$

(ii) For what temperatures would you believe the results. Explain.

↳ The Ising model converges at high temperatures which is why I would believe my results for $T = 10$.

(iii) What is $\lim_{T\to 0} \theta_T$ and $\lim_{T\to\infty} \theta_T$?

↳ $\lim_{T\to 0} \theta_T$ →

$\Rightarrow P(\vec{x}) = \frac{1}{Z_T} \exp\left(\frac{1}{T} \sum_{\{(i,j),(k,l)\}} x_{ij} x_{kl}\right)$

$\Rightarrow$ ratio of two summations $\varepsilon_1, \varepsilon_2$ → $\exp\left(\frac{1}{T}(\varepsilon_1 - \varepsilon_2)\right) = \begin{cases} 0 & \varepsilon_1 < \varepsilon_2 \\ \infty & \varepsilon_1 > \varepsilon_2 \end{cases}$

$\Rightarrow$ max of summation is achieved when →

$P(x_{ij} = 1) = P(x_{ij} = -1) \simeq 1/2$

$\therefore \lim_{T\to 0} \theta_T = E_T = \frac{1}{2} e^{1+1} + \frac{1}{2} e^{-1-1} = \frac{e^2}{2} + \frac{e^{-2}}{2}$

↳ $\lim_{T\to\infty} \theta_T$ →

$\Rightarrow P(\vec{x}) = \frac{1}{Z_T} \exp\left(\frac{1}{T} \sum_{\{(i,j),(k,l)\}} x_{ij} x_{kl}\right)$

$\Rightarrow$ almost every configuration has the same probability

$\Rightarrow$ X_{ij} are almost independent and $P(x_{ij} = \pm 1) = 1/2$

$\Rightarrow (x_{11}, x_{22}) \rightarrow \underbrace{(1,1)}_{1/4}, \underbrace{(1,-1), (-1,1)}_{1/2}, \underbrace{(-1,-1)}_{1/4}$

$\therefore \lim_{T\to 0} \theta_T = E_T = \frac{1}{4} e^{1+1} + \frac{1}{2} e^{0} + \frac{1}{4} e^{-1-1} = \frac{e^2}{4} + \frac{e^{-2}}{4} + \frac{1}{2}$

HOMEWORK 8

Anvita Bhagavathula

1. Suppose that I flip two fair coins, and let X_1 and X_2 be the indicator of heads on each coin, respectively. If $X_2 = 1$, then I flip a fair coin 3 times. If $X_2 = 0$, then I flip an unfair coin (probability 0.6 of heads) 3 times. Let X_3, X_4, X_5 be the indicator of heads for those three coin flips, respectively. Finally, if $X_1 + X_4 = 2$, then $X_6 = 1$, otherwise X_6 is the indicator of another fair coin flip.

   (a) What graph does (X_1, X_2) respect? Your answer should have as few edges as possible.

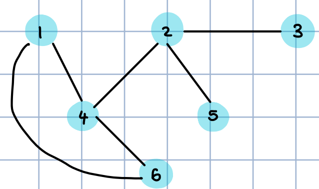
   (b) What graph does $(X_1, \ldots, X_6)$ respect? Your answer should have as few edges as possible.

   (c) Suppose $(Y_1, \ldots, Y_6)$ respects the graph from part b (with each X_i replaced by Y_i). Must (Y_1, Y_2) also respect the graph from part a (again with X_i replaced by Y_i)? Why or why not?

1. (a) Graph that (x_1, x_2) respect →

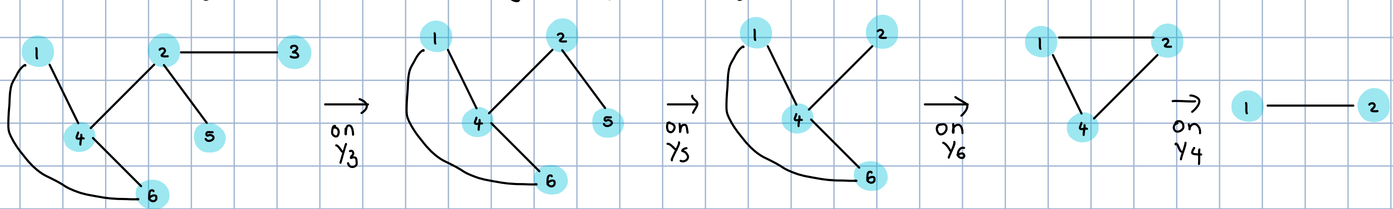
(x_1, x_2) are independent which is why there are no edges between them.

(b) Graph that $(x_1, \ldots, x_6)$ respect →



(X_3, X_4, X_5) are dependent on the outcome of x_2
X_6 is dependent on (x_1, x_4)
The above are connected via edges

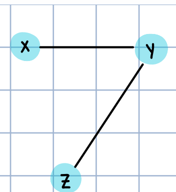
(c) Does (y_1, y_2) respect the graph in part (a) →
⇒ $(y_1, \ldots, y_6)$ respects the graph in part (b)
⇒ (y_1, y_2) will respect the graph in part (b) marginalised on y_3, y_4, y_5, y_6
⇒ this will give us the following (marginalised graphs) →



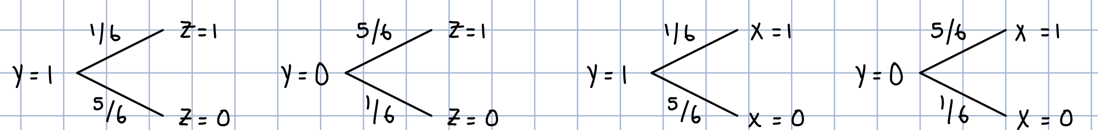
∴ we see that (y_1, y_2) does not respect the graph in part (a)

2. Let X, Y, Z be $\{0,1\}$-valued random variables whose joint distribution respects the graph with only an (X, Y) edge and a (Y, Z) edge. Give an example of such a distribution that is strictly positive everywhere and for which X and Z are dependent.

$X, Y, Z \sim \{0,1\}$ valued r.vs



⇒ $P(Y=1) = P(Y=0) = 1/2$



⇒ $P(Z=1) = P(Z=1 \mid Y=1) \cdot P(Y=1) + P(Z=1 \mid Y=0) \cdot P(Y=0)$
$= (1/6)(1/2) + (5/6)(1/2) = 1/2$

⇒ $P(Z=0) = P(Z=0 \mid Y=1) \cdot P(Y=1) + P(Z=0 \mid Y=0) \cdot P(Y=0)$
$= (5/6)(1/2) + (1/6)(1/2) = 1/2$

⇒ $P(X=1) = P(X=1 \mid Y=1) \cdot P(Y=1) + P(X=1 \mid Y=0) \cdot P(Y=0) = 1/2$
⇒ $P(X=0) = P(X=0 \mid Y=1) \cdot P(Y=1) + P(X=0 \mid Y=0) \cdot P(Y=0) = 1/2$

the distribution is strictly positive

⇒ to show that X, Z are dependent, we need to show that $P(X=x, Z=z) \neq P(X=x)P(Z=z)$
⇒ $P(X=0, Z=0) = P(X=0, Z=0 \mid Y=0)\, P(Y=0) + P(X=0, Z=0 \mid Y=1)\, P(Y=1)$
$= (1/6)(1/6)(1/2) + (5/6)(5/6)(1/2) = 13/36$
⇒ $P(X=0)\, P(Z=0) = (1/2)(1/2) = 1/4$

∴ $P(X=0, Z=0) \neq P(X=0)\, P(Z=0)$ and X, Z are dependent

3. Suppose $X = (X_1, \ldots, X_7)$ respects the graph G with vertices labeled $1, \ldots, 7$, respectively, and with edges $(1,2), (2,3), (3,4), (4,5), (5,6), (6,7), (1,4), (4,6), (2,7)$.

(a) Must X_4 and X_7 be conditionally independent given X_6? Explain your answer.

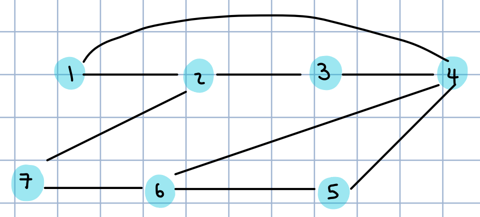
(b) Must X_4 and X_7 be conditionally independent given X_2 and X_6? Explain your answer.

(c) What is the simplest graph that $X_2, \ldots, X_7$ must respect?

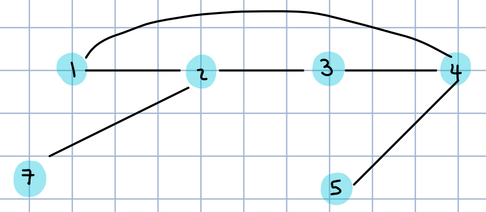
(d) What is the simplest graph that $X_3, \ldots, X_7$ must respect?

(e) What is the simplest graph that $X_4, \ldots, X_7$ must respect?

(f) Consider the conditional distribution of (X_2, X_4) given $X_1 = x_1$ and $X_3 = x_3$. What graph does this distribution respect? Your answer should have as few edges as possible.

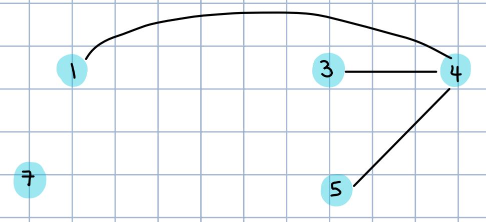


(a) (X_4, X_7) independence when conditioning on $X_6 \rightarrow$



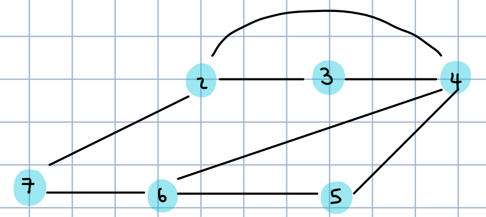
(X_4, X_7) are not independent for the graph conditioned on X_6 as there is still a path that exists between them

(b) (X_4, X_7) independence when conditioning on $X_2, X_6 \rightarrow$



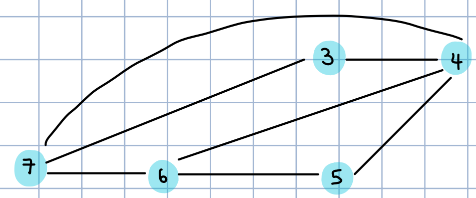
(X_4, X_7) are independent for the graph conditioned on X_6 and X_2 as there is no edge that connects the two nodes /no path between them.

(c) Simplest graph that $X_2, \ldots, X_7$ respect $\rightarrow$



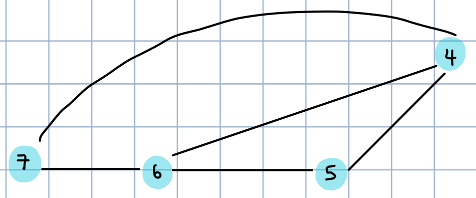
graph marginalised on X_1

(d) Simplest graph that $X_3, \ldots, X_7$ respect $\rightarrow$



graph marginalised on X_1, X_2

(e) Simplest graph that $X_4, \dots, X_7$ respect →



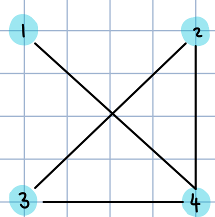
graph marginalised on X_1, X_2, X_3

(f) Conditional distribution of (X_2, X_4) given $X_1 = x_1$ and $X_3 = x_3$ →

Here we have conditioned on X_1, X_3 and marginalised on X_1, X_5, X_6, X_7.

4. Let $X = (X_1, X_2, X_3, X_4)$. Draw a graph G with as many edges as possible such that whenever X respects G we can conclude that both

- X_1 and X_2 are conditionally independent given X_3 and X_4; and
- X_1 and X_3 are conditionally independent given X_2 and X_4.



↳ conditioning on X_3, X_4 →

X_1, X_2 are independent

↳ conditioning on X_2, X_4 →

X_1, X_3 are independent

5. Consider the following joint pmf for the random variables $X_1, \dots, X_6$ over $\mathbb{Z}^6$.

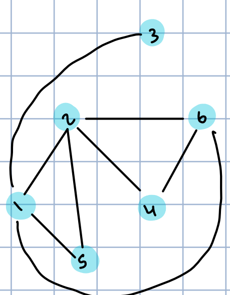
$$p(x_1, \dots, x_6) \triangleq \beta (x_1 x_2)^2 (\sin(x_2 + x_5))^4 (\cos(x_1 e^{x_3}))^2 (x_2 + x_4 + x_6)^{10} e^{-\sum_{i=1}^{6} x_i^2 - |x_1 x_6| - |x_1 x_5|}$$

where β is a normalization constant. Prove that X_4 and X_6 are conditionally independent from X_3 and X_5 given X_1 and X_2. (Graphical proofs, with explanation, are acceptable.)

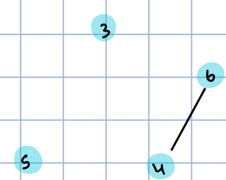
$$\Rightarrow p(x_1, \dots, x_6) = \beta (x_1 x_2)^2 (\sin(x_2 + x_5))^4 (\cos(x_1 e^{x_3}))^2 (x_2 + x_4 + x_6)^{10} e^{-\sum_{i=1}^{6} x_i^2 - |x_1 x_6| - |x_1 x_5|}$$

$$= \beta (x_1 x_2)^2 (\sin(x_2 + x_5))^4 (\cos(x_1 e^{x_3}))^2 (x_2 + x_4 + x_6)^{10} e^{-x_1^2} e^{-x_2^2} \dots e^{-x_6^2} e^{-|x_1 x_6|} e^{-|x_1 x_5|}$$

⇒ from the above we can obtain the following graph →



$\xrightarrow{\text{conditioning on } X_1, X_2}$



∴ after conditioning on X_1, X_2 we see that X_4, X_6 are independent from X_3 and X_5.

HOMEWORK 9

Anvita Bhagavathula

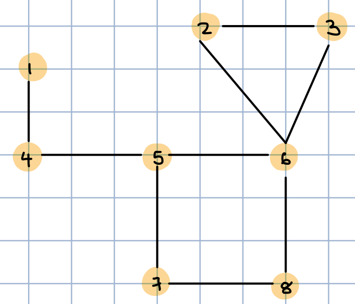
1. Consider the following pmf for $X = (X_1, \dots, X_8)$:

$$p_X(x) = K \frac{x_2 x_6 (x_4 x_7)^{x_5}}{(1 + |x_1 - x_4|)\Gamma(1+x_5)} \exp(-x_2 x_3 x_6 - (x_7 - x_8)^2 - (x_5 - x_6)^2 - (x_6 - x_8)^2)$$

where $x = (x_1, \dots, x_8)$, where each $x_i \in \{0.01, 0.02, 0.03, \dots, 1\} = \{j/100 : j = 1, \dots, 100\}$ and where K is a normalisation constant chosen so that the pmf sums to 1. $\Gamma(a)$ is the gamma function.

(a) What graph does X respect?

$\Rightarrow p_X(x) \propto \varphi(x_2)\varphi(x_6)\varphi(x_4, x_5)\varphi(x_7, x_5)\varphi(x_1, x_4)\varphi(x_5)\varphi(x_2, x_3, x_6)\varphi(x_7, x_8)\varphi(x_5, x_6)\varphi(x_6, x_8)$



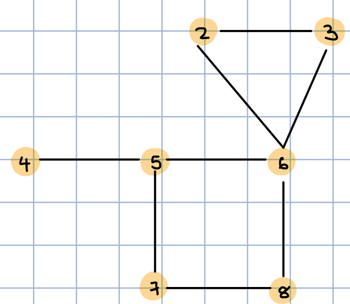
(b) Design a visitation schedule for computing K and $\arg\max_x p_X(x)$ via dynamic programming and indicate the workload in the big-O format.

↳ the visitation schedule should try to minimise the number of neighbours of each marginalised node in the dynamic programming algorithm.

↳ we choose this visitation schedule →

$x_1 \to x_4 \to x_5 \to x_7 \to x_8 \to x_6 \to x_2 \to x_3$

① marginalise on x_1

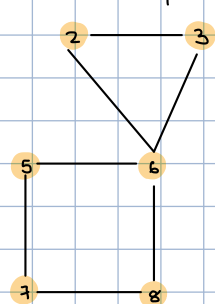


$$h_1(x_4) = \sum_{x_4} \sum_{x_1} \varphi(x_1, x_4)$$

workload $= O(d \cdot |S|^{1+1}) = O(d \cdot |S|^2)$

↳ all workloads will be for $d = 8$

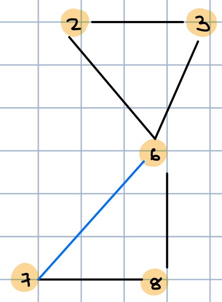
② marginalise on x_4



$$h_4(x_5) = \sum_{x_5} \sum_{x_4} h_1(x_4)\varphi(x_4, x_5)$$

workload $= O(d \cdot |S|^{1+1}) = O(d \cdot |S|^2)$

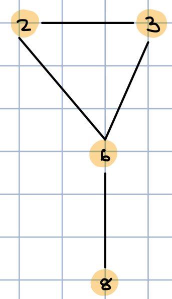
③ marginalise on x_5



$$h_5(x_6,x_7) = \sum_{x_6}\sum_{x_7}\sum_{x_5} h_4(x_5)\,\varphi(x_5)\,\varphi(x_5,x_7)\,\varphi(x_5,x_6)$$

workload $= O(d\cdot|S|^{2+1}) = O(d\cdot|S|^3)$

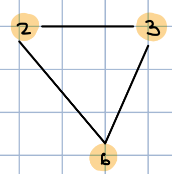
④ marginalise on x_7



$$h_7(x_6,x_8) = \sum_{x_6}\sum_{x_8}\sum_{x_7} h_5(x_6,x_7)\,\varphi(x_7,x_8)$$

workload $= O(d\cdot|S|^{2+1}) = O(d\cdot|S|^3)$

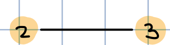
④ marginalise on x_8



$$h_8(x_6) = \sum_{x_6}\sum_{x_8} h_7(x_6,x_8)\,\varphi(x_6,x_8)$$

workload $= O(d\cdot|S|^{1+1}) = O(d\cdot|S|^2)$

⑤ marginalise on x_6



$$h_6(x_2,x_3) = \sum_{x_3}\sum_{x_2}\sum_{x_6} h_8(x_6)\,\varphi(x_2,x_3,x_6)\,\varphi(x_6)$$

workload $= O(d\cdot|S|^{2+1}) = O(d\cdot|S|^3)$

⑤ marginalise on x_2

$$h_2(x_3) = \sum_{x_3}\sum_{x_2} h_6(x_2,x_3)\,\varphi(x_2)$$

workload $= O(d\cdot|S|^{1+1}) = O(d\cdot|S|^2)$

=> Z = normalisation constant = $\sum h_2(x_3)$

$\therefore$ Big-O workload (overall) $= O(8\cdot|S|^3)$

$$K = \frac{1}{\sum h_2(x_3)}$$

(c) Use MATLAB to find K and $\text{argmax}_x P(x)$ →
↳ see attached code

=> $K = 1.8206 \times 10^{-15}$

=> $\text{argmax}_x P(x) = [\,1 \quad 1 \quad 0.01 \quad 1 \quad 0.84 \quad 1 \quad 1 \quad 1\,]$