Applied Stochastic Analyis
Homework 02

Author:	Jaime Cascante-Vega
Email:	jc12343@nyu.edu
University ID:	N19711762

Problem 1

Use Metropolis-Hasting to construct a Markov chain with the standard Gaussian distribution in d-dimensions as its invariant distribution. Use your imagination (and lots of plots) to demonstrate that your code is correctly sampling the Gaussian. As your proposal distribution use $q(y|x) = N(x, s^2 I)$ and experiment with different values of s (I is the $d \times d$ identity matrix). Does it seem to converge faster for some values of s than for others? What happens when you increase d? Finally, try proposing moves in one dimension at a time, i.e. add a small Gaussian perturbation to just one coordinate, then accept/reject, then perturb another coordinate, accept/reject, etc. Does this work better or worse than perturbing all of the coordinates at once?

Solution

In Figures 1, 2 and 4 I present the value of the Markov chain for a 1D, 2D and 3-Dimensional (3D) search respectively. The value of s was varied for all dimensional search in log space, $s \in \{10, 1, 0.1, 0.01\}$. In Figure 3 I present the evolving Markov chains in the state space, also varying s. For all the dimension the decreasing the value of s, that is the Gaussian perturbation, obtained from the proposal, results in a slower convergence of the Markov chain to the true value (pink line). For $s = 0.01$ the chain did not even recover the true value estimate. Examination of the evolution in the search space, see Figure 3, present a cleaner version of the effect of s. For big values of s the steps are larger compared to smaller values of s. However, for $s \leq 0.1$ the chain is not advancing, at it looks it only advance in one of the dimensions. Increasing the number the dimension results in consequence in a biased exploration of the dimension that is close to the target distribution $\pi = \mathrm{N}(\mathbf{0}, I)$.

In Figures 5 and 7 I present the Markov chain proposing moves only in one dimension at a time for 2D and 3D search spaces. The algorithm reaches the true value within the first 10 iterations, basically after updating each dimension. The convergence can be easily visualized in evolution in the search space presented in Figure 6. All chains, and regardless the value of s reach the true value within few steps, thus achieving perfect convergence. As a consequence, perturbing each coordinate independently works much better than perturbing all at once.

1 dimensional (1D) Metropolis Hasting (MH)

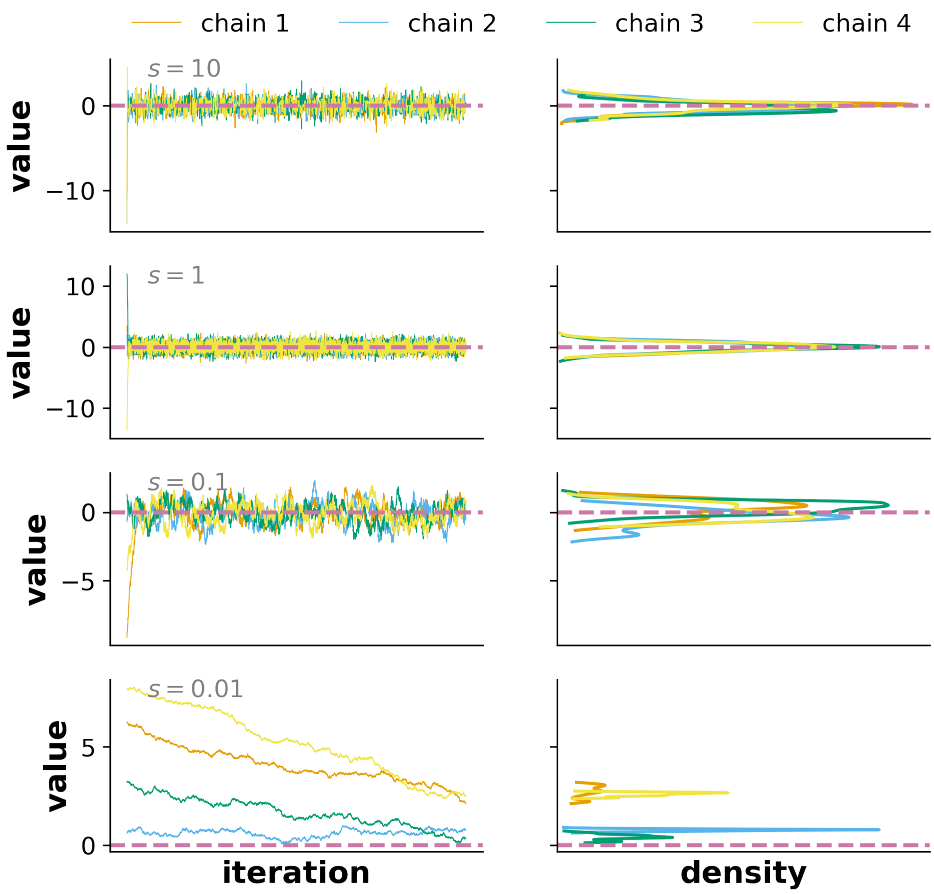

Figure 1: **1D Metropolis Hasting (MH) Markov Chain Monte Carlo (MCMC).** Left plots present the value of the 1 dimensional Markov Chain across 10000 iterations. I present 4 chains color-coded as indicated in the legend. Right plots present the estimated density, computed from random sampling the last 1000 iteration of the chain. Color match the estimated density for each chain. True value is presented a pink dashed line in all subplots. The value of s increases from upper to lower rows and is indicated in the upper left region of the left subplots.

2D MH

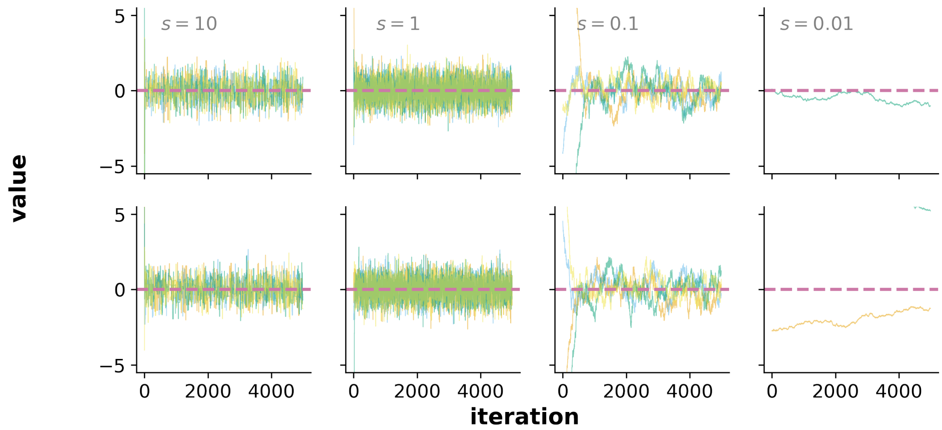

Figure 2: **2D MH-MCMC.** Upper and lower plots presents value of each parameter across 5000 Markov Chain iterations. I present 4 chains color-coded as indicated in the legend. True value is presented a pink dashed line in all subplots. The value of s increases from left to right column and is indicated in the upper left region of the left subplots.

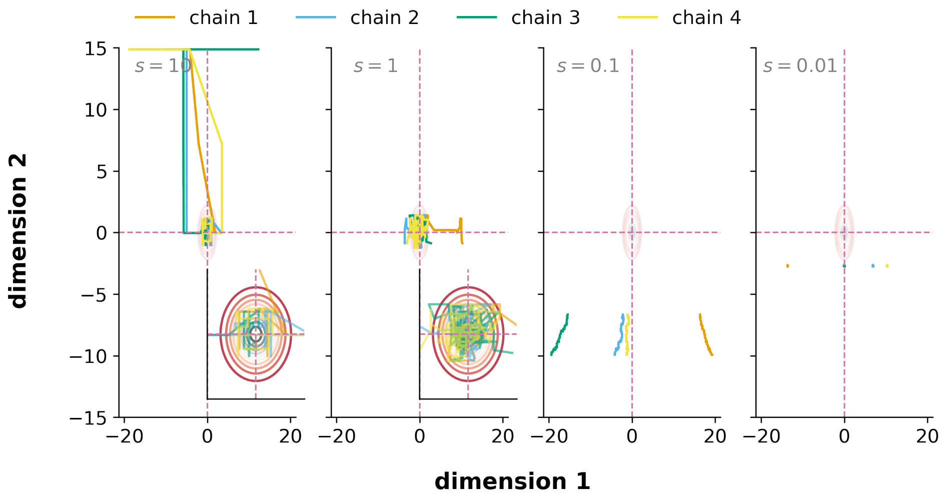

Figure 3: **Markov chain convergence in search space.** Evolution of 4 Markov chains as indicated in the legend. The target distribution $\mathrm{N}(\mathbf{0}, \mathbf{I})$ is presented with contour plots. Darker red value show level curves with less probability than light red values. For $s = 10$ and $s = 1$ I additionally present subplot insets to show the convergence of the chain to the target distribution. For $s = 0.1$ and $s = 0.01$ the chain do not reach the target distribution. For $s = 0.01$ the movement in the search space is negligible.

3D MH

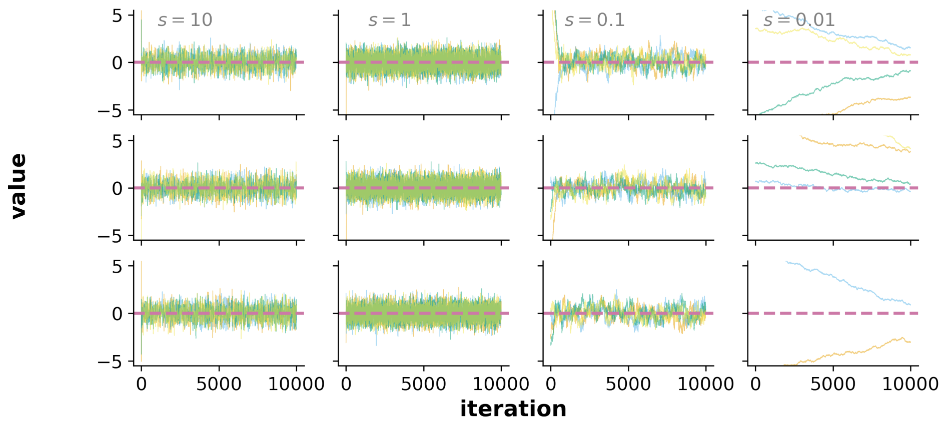

Figure 4: **3D MH-MCMC.** Upper, middle and lower plots presents value of each parameter across 10000 Markov Chain iterations. I present 4 chains color-coded. True value is presented a pink dashed line in all subplots. The value of s increases from left to right column and is indicated in the upper left region of the plots.

2D search, one dimension perturbation at a time

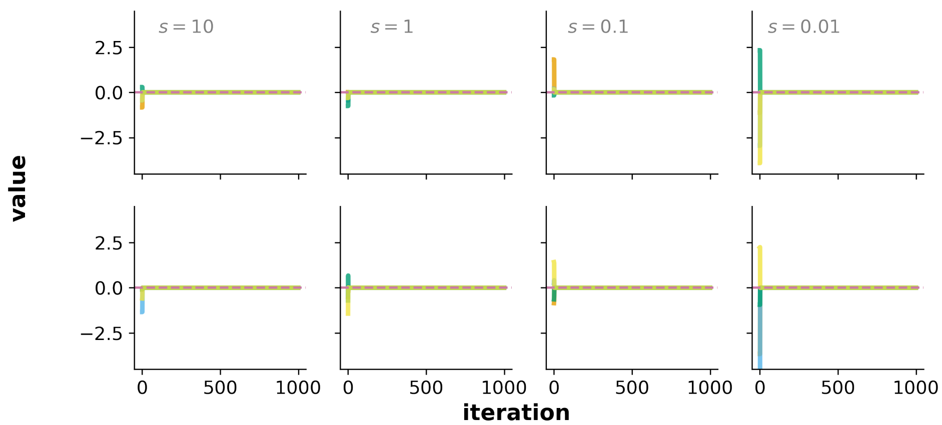

Figure 5: **2D MCMC updating one dimension at a time.** Upper and lower plots show the evolution of each parameters during the evolution of the Markov chain. I ran 4 different chains that are color coded. The true value is presented as a dashed pink line. The value of s increase from left to right.

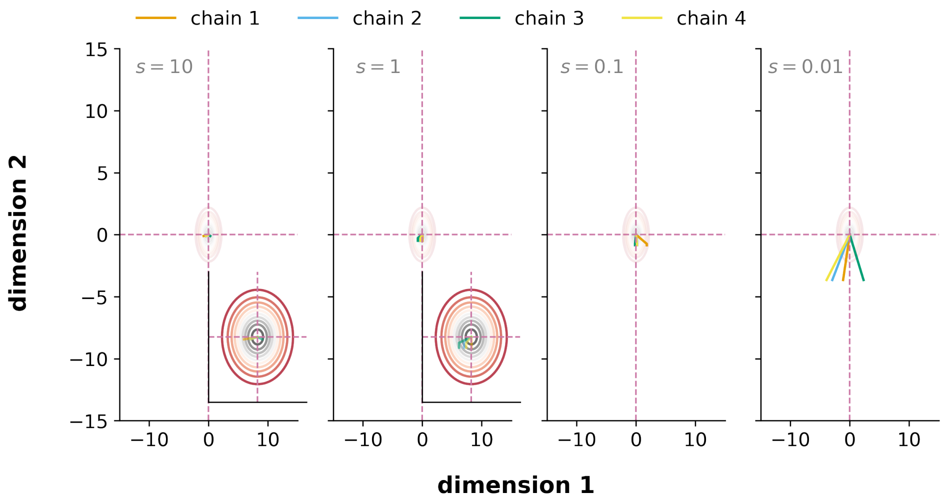

Figure 6: **Markov chain convergence in search space.** Evolution of 4 Markov chains as indicated in the legend. The target distribution $\mathrm{N}(\mathbf{0}, \mathbf{I})$ is presented with contour plots. Darker red value show level curves with less probability than light red values. For all the values of s the Markov chain convergence within a few iterations.

3D search, one dimension perturbation at a time

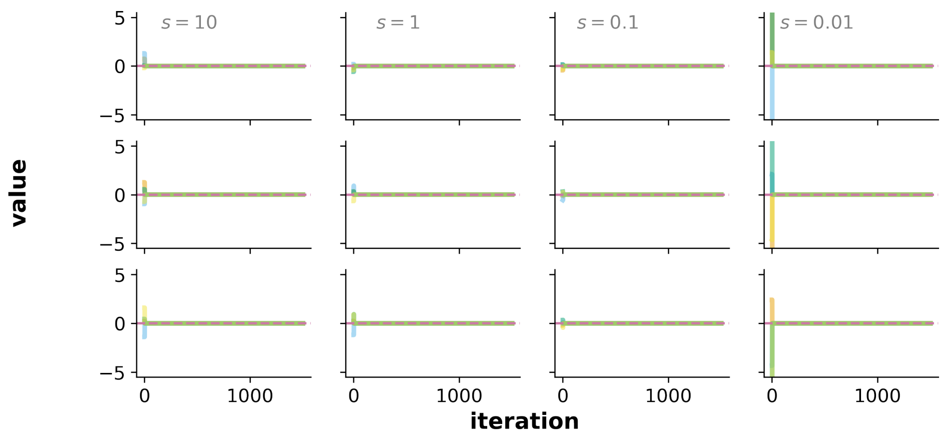

Figure 7: **3D MCMC updating one dimension at a time.** Upper, middle and lower plots presents value of each parameter across 2000 Markov Chain iterations. I present 4 chains color-coded. True value is presented a pink dashed line in all subplots. The value of s increases from left to right column and is indicated in the upper left region of the plots. All chain convergence within few iterations to the true value.

Problem 2

From the "Chapter 4" resource complete exercise 30.

Show that the eigenvalues of the transition matrix T for the process in the last example are the L roots of unity $e^{i2\pi/L}$ for $l = 0, \dots, L-1$. What are the eigenvalues of the matrix T^k for any k? What are the eigenvalues of the matrix

$$F = \frac{1}{L}\sum_{k=1}^{L} T^k?$$

Example 16. (last example) Consider sampling the uniform measure on $\mathbb{Z}_L$ by the Markov chain X^k with

$$P_i\left[X^1 = (i+1) \mod L\right] = 1$$

Solution.

The Transition matrix will be of size $L \times L$ and it's presented below

$$T = \begin{pmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & & 0 \\ 0 & 0 & 0 & 1 & \ddots & 0 \\ 0 & 0 & 0 & 0 & & 0 \\ \vdots & & \ddots & & \ddots & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

The characteristic polynomial will in general be

$$\det(T - \lambda I) = (-1)^L \cdot \left(1 - \lambda^L\right)$$

Replacing $\lambda = e^{i2\pi \cdot l/L}$ we have.

$$\begin{aligned} \det(T - \lambda I) =& (-1)^L \cdot \left(1 - (e^{i2\pi \cdot l/L})^L\right) \\ \det(T - \lambda I) =& (-1)^L \cdot \left(1 - e^{i2\pi \cdot l}\right) \end{aligned}$$

And by Euler's identity $e^{i\pi} = -1$ or $e^{i2\pi} = 1$. In consequence the eigenvalues are $\lambda = e^{i2\pi l/L}$.

What are the eigenvalues of the matrix T^k for any k?

The eigenvalues satisfy $A \cdot x = \lambda \cdot x$, multiplying by A, $A^2 \cdot x = \lambda \cdot A \cdot x$, and replacing $A \cdot x$

$$
\begin{aligned}
Ax =& \lambda x \\
& \text{multiplying by } A \\
A^2x =& \lambda Ax \\
& \text{replacing } Ax \text{ by } \lambda x \\
A^2x =& \lambda^2 x
\end{aligned}
$$

The eigenvalues of A^2 are in fact λ^2. And using a similar argument, multiplying by A repeatedly, the eigenvalues of A^k are λ^k. The eigenvalues of T^k are shown in equation below.

$$\lambda_{A^k} = (\lambda_A)^k = \left(e^{i2\pi l/L}\right)^k = e^{i2\pi kl/L}$$

What are the eigenvalues of the matrix

$$F = \frac{1}{L}\sum_{k=1}^{L} T^k?$$

The eigenvalues of F satisfy $Fx = \lambda_F x$, and we know that $T^k x = \lambda_T^k x$, where λ_T are the eigenvalues of the transition matrix T. Expanding the sum we have

$$
\begin{aligned}
Fx =& \frac{1}{L}\left(T + T^2 + T^3 + \cdots + T^L\right)x \\
Fx =& \frac{1}{L}\left(Tx + T^2x + T^3x + \cdots + T^Lx\right) \\
Fx =& \frac{1}{L}\left(\lambda_T x + \lambda_T^2 x + \lambda_T^3 x + \cdots + \lambda_T^L x\right) \\
& \text{using } Fx = \lambda_F x \\
\lambda_F x =& \frac{1}{L}\left(\lambda_T x + \lambda_T^2 x + \lambda_T^3 x + \cdots + \lambda_T^K x\right) \\
& \text{so it's easy to see that} \\
\lambda_F =& \frac{1}{L}\left(\lambda_T + \lambda_T^2 + \lambda_T^3 + \cdots + \lambda_T^L\right) \\
\lambda_F =& \frac{1}{L}\sum_{k=1}^{L}\lambda_T^k \\
& \text{replacing } \lambda_T^k = e^{i2\pi kl/L} \\
\lambda_F =& \frac{1}{L}\sum_{k=1}^{L} e^{i2\pi kl/L} \\
\lambda_F =& \frac{1}{L}\sum_{k=1}^{L}\left(e^{i2\pi}\right)^{kl/L} \\
\lambda_F =& \frac{1}{L}\sum_{k=1}^{L}(1)^{kl/L}
\end{aligned}
$$

The eigenvalues of F are 1. So one could similarly use the same eigenvalues of T, $\lambda = e^{i2\pi l/L}$.

Problem 3

From ASA, problem 3.1
Write down the transition probability matrix P of Ehrenfest's diffusion model and analyze its invariant distribution.

The Ehrenfest's diffusion model simulate exchange of gas molecules between two containers, A and B. Let X^n be the number of gas molecules in container A. $P(X^n = i) \quad i \in 0,1,2,\ldots,B$ is the probability that there are i balls at time n. Let's assume that $P(X^0 = B) = 1$, that is all the gas molecules are in container A at time $n = 0$. Assuming that in each time step only one molecule can move between containers the transition matrix is of size $(B+1)\times(B+1)$ and is presented below.

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & \ldots & 0 & 0 & 0 \\ \frac{1}{B} & 0 & \frac{B-1}{B} & 0 & \ldots & 0 & 0 & 0 \\ 0 & \frac{2}{B} & 0 & \frac{B-2}{B} & \ldots & 0 & 0 & 0 \\ 0 & 0 & \frac{3}{B} & 0 & \ldots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & 0 & \frac{B-(B-2)}{B} & 0 \\ 0 & 0 & 0 & 0 & \ldots & \frac{B-1}{B} & 0 & \frac{1}{B} \\ 0 & 0 & 0 & 0 & \ldots & 0 & 1 & 0 \end{pmatrix} \tag{1}$$

The invariant distribution $\boldsymbol{\pi}$ satisfies $\boldsymbol{\pi} = \boldsymbol{\pi}P$. Where $\boldsymbol{\pi} = [\pi_0, \pi_1, \ldots, \pi_B]$ and π_i is the stationary probability that there are i balls in the container A.

The invariant probability for π_i can be written as

$$\pi_j = \sum_{i=0}^{k} \pi_i P_{ij} \tag{2}$$

As in each row of P that are not the first or the last one the position P_{ij} there are only two non-zero terms. The probability P_{ij} will be non-zero only if $j = i \pm 1$, otherwise $P_{ij} = 0$. As a consequence Equation 2 will only have two non-zero terms. In expression below we write a general expression for x_j except for π_0 and π_B, $j \in [1, B-1]$. To compute π_j there must be calculated a normalization factor $\pi_j = \frac{x_j}{\sum_{i=0}^{B} x_i}$.

$$
\begin{aligned}
x_j =& x_{i-1}P_{j-1,j} + x_{j+1}P_{j+1,j} \\
& \text{and in general} \\
P_{j+1,j} =& \frac{j+1}{B} \\
P_{j-1,j} =& \frac{B-(j-1)}{B}
\end{aligned} \tag{3}
$$

I assumed all balls were initially in the container A or $\pi_0 = x_0 = 1$. And also $x_0 = x_1 \cdot \frac{1}{B}$. Solving for $x_1 = x_0 \cdot B = B$.
The expression for x_1 using the equation presented previously is shown below.

$$
\begin{aligned}
x_1 =& x_0 \cdot P_{0,1} + x_2 \cdot P_{2,1} \\
P_{0,1} =& \frac{B}{B} \\
P_{2,1} =& \frac{3}{B} \\
x_1 =& x_0 \cdot \frac{B}{B} + x_2 \cdot \frac{2}{B} \\
& \text{solving for } x_2, \text{using } x_1 = B,\, x_0 = 1 \\
x_2 =& \frac{B^2 - B}{2} \\
x_2 =& \frac{B(B-1)}{2}
\end{aligned}
$$

It's also of importance to note that $x_2 = x_1 \frac{(B-1)}{2}$.

Similarly, solving Equation 3 for x_{i+1} to continue applying the recurrence.

$$
\begin{aligned}
x_{i+1} =& x_i \frac{1}{P_{i+1,i}} - x_{i-1} \frac{P_{i-1,i}}{P_{i+1,i}} \\
x_{i+1} =& x_i \frac{B}{i+1} - x_{i-1} \frac{B-(i-1)}{i+1} \\
x_{i+1} =& x_i \frac{B}{i+1} - x_{i-1} \frac{B-(i-1)}{i+1}
\end{aligned} \tag{4}
$$

Solving for x_3, using $x_1 = B$ and $x_2 = \frac{B(B-1)}{2}$.

$$
\begin{aligned}
x_3 =& \frac{B(B-1)}{2} \frac{B}{3} - B \frac{B-1}{3} \\
x_3 =& \frac{B^2(B-1) - 2B(B-1)}{6} \\
x_3 =& \frac{B(B-1)(B-2)}{6}
\end{aligned} \tag{5}
$$

Also note that $x_3 = x_2 \dfrac{B-2}{3}$, and $x_2 = x_1 \dfrac{B-1}{2}$.
By induction one should expect that $x_4 = x_3 \dfrac{B-3}{3}$ and in general $x_{i+1} = x_i \dfrac{B-(i-1)}{i}$. However in the expression below I use equation 4 to solve for x_4.

$$
\begin{aligned}
x_{i+1} =& x_i \frac{1}{P_{i+1,i}} - x_{i-1} \frac{P_{i-1,i}}{P_{i+1,i}} \\
x_4 =& x_3 \frac{1}{P_{4,3}} - x_2 \frac{P_{2,3}}{P_{4,3}} \\
P_{4,3} =& \frac{4}{B}, \quad P_{2,3} = \frac{B-2}{B} \\
x_4 =& \frac{B(B-1)(B-2)}{6} \frac{B}{4} - \frac{3B(B-1)}{6} \frac{(B-2)}{4} \\
x_4 =& \frac{B(B-1)(B-2)(B-3)}{24}
\end{aligned} \tag{6}
$$

Here note that it could be factorized in terms of $x_1 = B$, $x_2 = x_1(B-1)/2$ and $x_3 = x_2(B-2)/3$.

$$
\begin{aligned}
x_4 =& x_1 \frac{(B-1)(B-2)(B-3)}{24} \\
x_4 =& x_2 \frac{(B-2)(B-3)}{12} \\
x_4 =& x_3 \frac{(B-3)}{4}
\end{aligned} \tag{7}
$$

From this induction we can conclude two things first that as suggested $x_{i+1} = x_i \frac{B-(i-1)}{i}$. As a result the Markov chain is in detailed balance. Here we exactly have that the transition probabilities π_i and π_{i+1} satisfy the recursive relationship $\pi_{i+1} = \pi_i \frac{B-(i-1)}{i}$.
This recursive relationship satisfy:

$$\pi_i = \binom{B}{i} \pi_0$$

And π_0 is the normalization factor

$$\pi_0 = \frac{1}{\sum_{i=1}^{B} x_i} = \frac{1}{\sum_{i=1}^{B} \binom{B}{i}} = 2^{-B}$$

In consequence

$$\pi_i = \binom{B}{i} 2^{-B}$$

Problem 4

From ASA, problem 3.5

Show that the recurrence relation $\boldsymbol{\mu} = \boldsymbol{\mu}_{n-1}\mathbf{P}$ can be written as

$$\mu_{n,i} = \mu_{n-1,i}\left(1 - \sum_{j\in S, j\neq i} p_{ij}\right) + \sum_{j\in S, j\neq i} \mu_{n-1,j}p_{ji} \tag{8}$$

The first term on the right-hand side gives the total probability of not making a transition from state i, while the last term is the probability of transition from one of states $j \neq i$ to state i.

The expression for $\mu_{n,i}$, using $\boldsymbol{\mu} = \boldsymbol{\mu}_{n-1}\mathbf{P}$ is presented below. Where $|S|$ is the size of the state space. Where $\boldsymbol{\mu}_n = \left[\mu_{n,1}, \mu_{n,2}, \ldots, \mu_{n,|S|}\right]$.

$$\mu_{n,k} = \sum_{i\in S} \mu_{n-1,i} \cdot p_{ij} \tag{9}$$

The sum of the probabilities per row and column is 1. That is $\sum_{i\in S} p_{i,j} = \sum_{j\in S} p_{i,j} = 1$. In consequence the $i-th$ position, p_{ij} can be written as $p_{ij} = 1 - \sum_{j\in S, j\neq i} p_{ij}$. Equation 8 can be factorized so the $i-th$ element is removed from equation 9, and is computed as $\mu_{n-1,i}\left(1 - \sum_{j\in S, j\neq i} p_{ij}\right)$. This result in the expression presented in equation 8.

Problem 5

From ASA, problem 3.16
Consider an irreducible Markov chain $\{X_n\}_{n\in\mathbb{N}}$ on a finite state space S. Let $H \subset S$. Define the first passage time $T_H = \inf\{n | X_n \in H, n \in \mathbb{N}\}$ and

$$h_i = P^i (T_h < \infty), \quad i \in S.$$

Prove that $\mathbf{h} = (h_i)_{i\in S}$ satisfies the equation

$$(\mathbf{I} - \mathbf{P}) \cdot \mathbf{h} = 0 \quad \text{with boundary condition } h_i = 1, \quad i \in H$$

where $\mathbf{P}$ is the transition probability matrix.

$$\begin{aligned}
h_i =& P_i (T_H < \infty) \\
h_i =& \sum_{j\in S} P_i (T_H < \infty | X_1 = j) \\
h_i =& \sum_{j\in S} P_i (T_H < \infty | X_1 = j) \\
& \text{by Markov } P_i (T_H < \infty | X_1 = j) = h_j \\
h_i =& \sum_{j\in S} h_j P_i (X_1 = j) \\
h_i =& \sum_{j\in S} h_j \cdot P_{ij} \\
0 =& h_i - \sum_{j\in S} h_j \cdot P_{ij} \\
& \text{and in matrix form} \\
0 =& (\mathbf{I} - \mathbf{P}) \cdot \mathbf{h}
\end{aligned}$$