DS102 - Homework 1

> If you are submitting a handwritten version please make sure your answers are legible as you may lose points otherwise.
> Data science is a collaborative activity. While you may talk with other about the homeworks, we ask that you write your solutions individually. If you do discuss the homework with others please include their name in your submission.
> **Due by: 1:59pm, Tuesday 10th September, 2019**

1. (10 points) You're at the state fair, and you decide to play the following two-stage game:

   In stage 1, you flip two independent coins, with probability of landing heads $\mathbb{P}[\text{flip}_1 = H] = p_1$ and $\mathbb{P}[\text{flip}_2 = H] = p_2$, respectively. If both coins land tails, the game ends. Otherwise, you move on to stage two, and receive a randomly-defined reward R, where R depends on your previous flips from stage 1. Specifically, let random variable Y denote the number of heads you received between the two coin flips, then you receive a reward as a random draw from the distribution

   $$R|Y \sim Unif(0, Y)$$

   In the next three questions, show the steps to reach your answer and give the name of any properties/rules used. For example, part (a) is done for you.

   (a) (0 points) What is the probability that you make it to stage 2?
   Example solution:

   $$\begin{aligned}\mathbb{P}[\text{flip}_1 = H \cup \text{flip}_2 = H] &= 1 - \mathbb{P}[\text{flip}_1 = T \cap \text{flip}_2 = T] \\ &= 1 - \mathbb{P}[\text{flip}_1 = T] \cdot [flip_2 = T] \quad (\text{flip}_1, \text{flip}_2 \text{ independent}) \\ &= 1 - (1 - p_1) \cdot (1 - p_2)\end{aligned}$$

   (b) (3 points) What is the expected payoff of the game, $\mathbb{E}[R]$?
   hint: to reduce computation, use the law of total expectation (also known as the tower property). The law of total expectation states that $\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid Y]]$.

   (c) (3 points) Conditional on the event that the first flip lands tails ($\text{flip}_1 = T$), what is the probability that you make it to stage 2?

   (d) (4 points) What is the expected payoff of the game $\mathbb{E}[R \mid \text{flip}_1 = T]$, conditioned on the event that the first flip lands tails? Similarly, what is $\mathbb{E}[R \mid \text{flip}_1 = H]$?

2. (15 points) In this question we will revise crucial concepts from linear algebra. We will begin by recalling the definition of eigenvectors. We will then see how eigenvectors interact with an idea called *quadratic forms.* Finally we will use these ideas to visualize bivariate Gaussian distributions. Throughout this question assume that we are only working with real numbers, we do not deal with complex numbers here.

Recall that the eigenvalues of a matrix A is given by the set of all real numbers λ such that there exists a non-zero vector v (called an eigenvector) such that

$$Av = \lambda v.$$

Another definition that will be important throughout this question is the concept of a positive definite (PD) and positive semi-definite (PSD) matrix. A symmetric matrix is considered to be PD if all its eigenvalues are strictly positive, while a matrix is positive semi-definite if all its eigenvalues are greater than or equal to zero.

One way to think of PSD matrices is as the matrix "generalization" of a non-negative real number (and PD matrices as the matrix "generalization" of a strictly positive real number). In particular a PSD matrix A always has a square root, that is, there exists a matrix $\sqrt{A}$ such that $A = \sqrt{A}\sqrt{A}$. We can check this by diagonalizing A, which is always possible since it is a symmetric matrix (if you don't remember much about diagonalization you can check out this link). If we let D be the diagonal matrix of the eigenvalues of A, L be the matrix with the corresponding unit-length eigenvectors as its columns, and $\sqrt{D}$ be the elementwise square root of D then:

$$\begin{aligned} A &= LDL^{-1} \\ &= L\sqrt{D}\sqrt{D}L^{-1} \\ &= L\sqrt{D}I\sqrt{D}L^{-1} \\ &= L\sqrt{D}L^{-1}L\sqrt{D}L^{-1}. \end{aligned}$$

Hence $\sqrt{A} = L\sqrt{D}L^{-1}$, note that this square root is also symmetric since $L^{-1} = L^\top$ in this case.

(a) (5 points) Compute the eigenvalues and an associated non-zero eigenvector for the following matrices.

$$\begin{bmatrix} 4 & 0 \\ 0 & 16 \end{bmatrix} \qquad \begin{bmatrix} 1/2 & -1/2 \\ -1/2 & 1/2 \end{bmatrix} \qquad \begin{bmatrix} 17/8 & 15/8 \\ 15/8 & 17/8 \end{bmatrix}$$

(b) (5 points) We now investigate the behaviour of quadratic forms and their relationship with eigenvalues. The quadratic form of the matrix A is defined as the polynomial $x^\top Ax$. If A is PSD we can think of its quadratic form as the function from a vector x to the squared length of the vector $\sqrt{A}x$, since $x^\top Ax = x^\top \sqrt{A}^\top \sqrt{A}x = \left\|\sqrt{A}x\right\|_2^2$, where we use $\|x\|_2 = \sqrt{x^\top x}$.

- This exercise can be done in whatever your favourite plotting library is. For each of the three matrices defined in part (a), plot the level curves of their quadratic form at the values $c = 1, 2, 4, 8$ where $c = x^\top Ax$. Make sure that each level is clearly labelled.
- What are the lengths of the eigenvectors as they intersect the level curves?

- Is there anything noteworthy about these lengths?

(c) (5 points) Finally we turn to bivariate Gaussian distributions which are the most popular way to use quadratic forms. Recall that the multivariate Gaussian distribution of a random vector x with mean $\mu \in \mathbb{R}^n$ and covariance $\Sigma \in \mathbb{R}^{n \times n}$ is defined to be

$$\mathcal{N}(\mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right).$$

Note that the term inside the exponential is simply the quadratic form of Σ^{-1} centered at μ and multiplied by $-1/2$.

- This exercise can be done in whatever your favourite plotting library is. Plot the level curves of the bivariate Gaussians with $\mu = \begin{bmatrix} 1 & -1 \end{bmatrix}^\top$ and Σ^{-1} (**note the inverse here**) given by the three matrices defined in part (a) at the values $c = 0.001, 0.01, 0.05, 0.1$. Make sure that each level is clearly labelled. If Σ does not exist you do not have to plot the level curves for that matrix.
- How do the plots you generated in part (b) relate to those in part (c)?
- Give a probabilistic interpretation to the plots you generated in terms of the covariances, variances and the means of the two random variables in the vector x. In particular, how is the orientation and shape of the level curves influenced by these quantities? What about the length of the level curves?

3. (15 points) This question covers the topic of Markov chains, which we'll use later in the class. Depending on which classes you've taken before, students will have differing backgrounds, so we suggest reading or reviewing Chapter 10 in the Prob140 textbook (`http://prob140.org/textbook/chapters/Chapter_10/`) and/or Chapters 7.1-7.3 in the EECS126 textbook (Bertsekas and Tsitiklis, Introduction to Probability, 2009).

A Markov chain is defined by a set of *states* S, and a set of *transition probabilities* defined on pairs of states. For instance, the diagram in Figure 1 below has three states, a, b, and c. At each time t, X_t is a random variable representing the state of the chain. Denote p_{ij} as the *one-step transition probability* for any time t and any $i, j \in \{a, b, c\}$:

$$p_{ij} = \mathbb{P}[X_{t+1} = j | X_t = i]$$

Consider the state space and transition structure from Figure 1. We will denote transition probabilities in matrix form, where rows correspond to the "from" state, and columns correspond to the "to" state, so that the 3×3 *transition probability matrix* takes the form

$$\begin{bmatrix} p_{aa} & p_{ab} & p_{ac} \\ p_{ba} & p_{bb} & p_{bc} \\ p_{ca} & p_{cb} & p_{cc} \end{bmatrix}$$

We will explore the following properties of Markov chains:

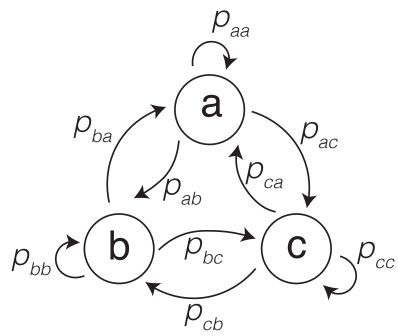

Figure 1: Markov chain structure for problem 3.

- A Markov chain is said to be *ergodic* (*irreducible*) if each state is reachable from every other state (with nonzero probability).
- Each state in the Markov chain is said to have *period* d if, starting at state i, the chain can come back to i only at times that are multiples of d. If all states have period 1, the Markov chain is said to be *aperiodic*.
- For ergodic and aperiodic Markov chains, there is a consistent notion of *steady state probabilities*

$$\pi(j) = \lim_{t\to\infty} \mathbb{P}[X_t = j | X_0 = i]$$

(if these are unfamiliar terms, consult the resources from the beginning of the problem.)

For each of the following three transition probability matrices and the state space from Figure 1, state (i) if the Markov chain is ergodic, (ii) if the chain is periodic or aperiodic, and (iii) give the steady state probabilities. If the Markov chain is not ergodic or is periodic, for part (iii) give the limiting behavior of the when initial distribution of mass on states $[a, b, c]$ $[1/3, 1/3, 1/3]$ for X_0.

(a) (3 points)

$$P_1 = \begin{bmatrix} 1/2 & 1/4 & 1/4 \\ 1/4 & 1/2 & 1/4 \\ 1/4 & 1/4 & 1/2 \end{bmatrix}$$

(b) (3 points)

$$P_2 = \begin{bmatrix} 2/3 & 1/3 & 0 \\ 1/2 & 1/2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

(c) (3 points)

$$P_3 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

(d) (6 points) The steady state probabilities, if they exist, can equivalently be written as solving the balance equations $\pi = \pi P$. These are called balance questions because for each state j, the proportion of mass flowing out of state j ($\pi(j)$) is equal to the proportion of mass flowing into state j ($\sum_i \pi(i) P_{ij}$.)

(i) For the Markov chain in Figure 2, write the balance equations for each steady state probability $\pi(j), j = 1, ..., n$ as a function of elements of the other steady state probabilities $\pi(i), i \neq j$, as well as the parameter c which controls amount of flow toward the right.

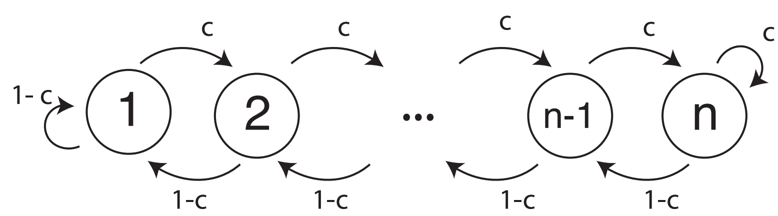

Figure 2: Markov chain structure for problem 3.

(ii) Use the balance equations to solve for the steady state probabilities $\pi(j)$ in terms of only c and the index j.

(iii) As a special case, what are the steady state probabilities when $c = 1/2$?