CS 70 Spring 2018 — Discrete Mathematics and Probability Theory — Babak Ayazifar and Satish Rao — HW 13

## Sundry

Before you start your homework, write down your team. Who else did you work with on this homework? List names and email addresses. (In case of homework party, you can also just describe the group.) How did you work on this homework? Working in groups of 3-5 will earn credit for your "Sundry" grade.

Please copy the following statement and sign next to it:

*I certify that all solutions are entirely in my words and that I have not looked at another student's solutions. I have credited all external sources in this write up.*

I certify that all solutions are entirely in my words and that I have not looked at another student's solutions. I have credited all external sources in this write up. (*signature here*)

# 1 LLSE and Graphs

Consider a graph with $n$ vertices numbered 1 through $n$, where $n$ is a positive integer $\geq 2$. For each pair of distinct vertices, we add an undirected edge between them independently with probability $p$. Let $D_1$ be the random variable representing the degree of vertex 1, and let $D_2$ be the random variable representing the degree of vertex 2.

(a) Compute $\mathbb{E}[D_1]$ and $\mathbb{E}[D_2]$.

(b) Compute $\text{var}(D_1)$.

(c) Compute $\text{cov}(D_1, D_2)$.

(d) Using the information from the first three parts, what is $L(D_2 \mid D_1)$?

**Solution:**

Throughout this problem, let $X_{i,j}$ be an indicator random variable for whether the edge between vertex $i$ and vertex $j$ exists, for $i, j = 1, \ldots, n$. Note that $X_{i,j} = X_{j,i}$.

(a) Observing that $D_1, D_2 \sim \text{Binomial}(n-1, p)$, we obtain $\mathbb{E}[D_1] = \mathbb{E}[D_2] = (n-1)p$.

Anyway, it is good to review how we derived the expectation of the binomial distribution in the first place. By linearity of expectation,

$$\mathbb{E}[D_1] = \mathbb{E}\left[\sum_{i=2}^{n} X_{1,j}\right] = \sum_{i=2}^{n} \mathbb{E}[X_{1,j}] = (n-1)\,\mathbb{E}[X_{i,j}] = (n-1)p.$$

By symmetry, $\mathbb{E}[D_2] = (n-1)p$ also.

(b) Since $D_1, D_2 \sim \text{Binomial}(n-1, p)$, then $\operatorname{var} D_1 = \operatorname{var} D_2 = (n-1)p(1-p)$.

Again, it is good to review how we calculated the variance of the binomial distribution.

Solution 1: Write the variance of $D_1$ as a sum of covariances.

$$\begin{aligned}\operatorname{var}(D_1) &= \operatorname{cov}\Big(\sum_{i=2}^{n} X_{1,i}, \sum_{i=2}^{n} X_{1,i}\Big) = (n-1)\operatorname{var}(X_{1,2}) + \big((n-1)^2 - (n-1)\big)\operatorname{cov}(X_{1,2}, X_{1,3}) \\ &= (n-1)p(1-p) + 0 = (n-1)p(1-p).\end{aligned}$$

We used the fact that $X_{1,i}$ and $X_{1,j}$ are independent if $i \neq j$, so their covariance is zero.

Solution 2: Compute the variance directly.

$$\begin{aligned}\operatorname{var}(X) &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}\Big[\Big(\sum_{i=2}^{n} X_{1,i}\Big)^2\Big] - (n-1)^2 p^2 \\ &= (n-1)\,\mathbb{E}[X_{1,2}^2] + \big((n-1)^2 - (n-1)\big)\,\mathbb{E}[X_{1,2}X_{1,3}] - (n-1)^2 p^2 \\ &= (n-1)p + (n^2 - 3n + 2)p^2 - (n-1)^2 p^2 \\ &= (n-1)p + (n-1)(n-2)p^2 - (n-1)^2 p^2 = (n-1)p\big(1 + (n-2)p - (n-1)p\big) \\ &= (n-1)p(1-p)\end{aligned}$$

(c) We can write

$$\operatorname{cov}(D_1, D_2) = \operatorname{cov}\Big(\sum_{i=2}^{n} X_{1,i}, \sum_{i=1, i\neq 2}^{n} X_{2,i}\Big) = \sum_{i=2}^{n} \sum_{j=1, j\neq 2}^{n} \operatorname{cov}(X_{1,i}, X_{2,j}).$$

Note that all pairs of $X_{1,i}, X_{2,j}$ are independent except for when $i = 2$ and $j = 1$, so all terms in the sum are zero except for $\operatorname{cov}(X_{1,2}, X_{2,1})$, and our covariance is just equal to $\operatorname{cov}(X_{1,2}, X_{2,1}) = \operatorname{var}(X_{1,2}) = p(1-p)$.

(d) Since

$$L(D_2 \mid D_1) = \mathbb{E}[D_2] + \frac{\operatorname{cov}(D_1, D_2)}{\operatorname{var}(D_1)}(D_1 - \mathbb{E}[D_1]),$$

we plug in our values from the first three parts to get that

$$\begin{aligned}L(D_2 \mid D_1) &= (n-1)p + \frac{p(1-p)}{(n-1)p(1-p)}\big(D_1 - (n-1)p\big) \\ &= (n-1)p + \frac{1}{n-1}\big(D_1 - (n-1)p\big) = \frac{1}{n-1}D_1 + (n-2)p.\end{aligned}$$

## 2 Swimsuit Season

In the swimsuit industry, it is well-known that there is a "swimsuit season". During this time, swimsuit sales skyrocket!

We will model this with a random variable $X$ which is either $\lambda_L$ or $\lambda_H$ with equal probability; $\lambda_L$ represents the mean number of customers in a day when swimsuits are not in season, and $\lambda_H$ represents the mean number of customers during swimsuit season. So, $\lambda_L$ is the "low rate" and $\lambda_H$ is the "high rate". The number of customer arrivals $Y$ on a particular day is modeled as a Poisson random variable with mean $X$.

You observe $Y$ customers on a certain day, and the task is to estimate $X$.

(a) What is $L[X \mid Y]$?

(b) What is $\mathbb{E}[X \mid Y]$?

**Solution:**

(a) The key idea here is that $X$ gives information about $Y$, so to calculate any quantity involving $Y$, it is helpful to condition on $X$.

First, we observe that since $X$ is the mean of $Y$, $\mathbb{E}[Y \mid X] = X$. So,

$$\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y \mid X]] = \mathbb{E}[X] = \frac{1}{2}(\lambda_L + \lambda_H).$$

Now, we compute $\operatorname{cov}(X,Y)$.

$$\begin{aligned}
\operatorname{cov}(X,Y) &= \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[XY \mid X]] - \mathbb{E}[X]^2 = \mathbb{E}[X\,\mathbb{E}[Y \mid X]] - \mathbb{E}[X]^2 \\
&= \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{1}{2}(\lambda_L^2 + \lambda_H^2) - \frac{1}{4}(\lambda_L + \lambda_H)^2 = \frac{1}{4}(\lambda_H^2 - 2\lambda_L\lambda_H + \lambda_L^2) \\
&= \frac{1}{4}(\lambda_H - \lambda_L)^2.
\end{aligned}$$

Next is $\operatorname{var}(Y)$.

$$\operatorname{var}(Y) = \mathbb{E}[Y^2] - \mathbb{E}[Y]^2 = \mathbb{E}[\mathbb{E}[Y^2 \mid X]] - \mathbb{E}[\mathbb{E}[Y \mid X]]^2 = \mathbb{E}[X^2 + X] - \mathbb{E}[X]^2$$

(remember that if $Y$ has the Poisson distribution with mean $\lambda$, then $\mathbb{E}[Y^2] = \lambda^2 + \lambda$)

$$= \frac{1}{2}(\lambda_L^2 + \lambda_H^2) + \frac{1}{2}(\lambda_L + \lambda_H) - \frac{1}{4}(\lambda_L + \lambda_H)^2 = \frac{1}{4}(\lambda_H - \lambda_L)^2 + \frac{1}{2}(\lambda_L + \lambda_H).$$

Hence,

$$\begin{aligned}
L[X \mid Y] &= \mathbb{E}[X] + \frac{\operatorname{cov}(X,Y)}{\operatorname{var}(Y)}(Y - \mathbb{E}[Y]) \\
&= \frac{1}{2}(\lambda_L + \lambda_H) + \frac{(\lambda_H - \lambda_L)^2/4}{(\lambda_H - \lambda_L)^2/4 + (\lambda_L + \lambda_H)/2}\left(Y - \frac{1}{2}(\lambda_L + \lambda_H)\right).
\end{aligned}$$

(b) We calculate $\mathbb{P}(X = \lambda_L \mid Y = y)$.

$$\mathbb{P}(X = \lambda_L \mid Y = y) = \frac{\mathbb{P}(Y = y \mid X = \lambda_L)\mathbb{P}(X = \lambda_L)}{\mathbb{P}(Y = y \mid X = \lambda_L)\mathbb{P}(X = \lambda_L) + \mathbb{P}(Y = y \mid X = \lambda_H)\mathbb{P}(X = \lambda_H)}$$
$$= \frac{e^{-\lambda_L}\lambda_L^y/y!}{e^{-\lambda_L}\lambda_L^y/y! + e^{-\lambda_H}\lambda_H^y/y!} = \frac{e^{-\lambda_L}\lambda_L^y}{e^{-\lambda_L}\lambda_L^y + e^{-\lambda_H}\lambda_H^y}.$$

Therefore,

$$\mathbb{E}[X \mid Y = y] = \lambda_L \cdot \frac{e^{-\lambda_L}\lambda_L^y}{e^{-\lambda_L}\lambda_L^y + e^{-\lambda_H}\lambda_H^y} + \lambda_H \cdot \frac{e^{-\lambda_H}\lambda_H^y}{e^{-\lambda_L}\lambda_L^y + e^{-\lambda_H}\lambda_H^y}$$

so

$$\mathbb{E}[X \mid Y] = \frac{e^{-\lambda_L}\lambda_L^{Y+1} + e^{-\lambda_H}\lambda_H^{Y+1}}{e^{-\lambda_L}\lambda_L^Y + e^{-\lambda_H}\lambda_H^Y}.$$

This is a very non-linear function of $Y$, which illustrates that in general, the MMSE does not equal the LLSE.

## 3 Quadratic Regression

In this question, we will find the best quadratic estimator of $Y$ given $X$. First, some notation: let $\mu_i$ be the $i$th moment of $X$, i.e. $\mu_i = \mathbb{E}[X^i]$. Also, define $\beta_1 = \mathbb{E}[XY]$ and $\beta_2 = \mathbb{E}[X^2Y]$. For simplicity, we will assume that $\mathbb{E}[X] = \mathbb{E}[Y] = 0$ and $\mathbb{E}[X^2] = \mathbb{E}[Y^2] = 1$. (Note that this poses no loss of generality, because we can always transform the random variables by subtracting their means and dividing by their standard deviations.) We claim that the best quadratic estimator of $Y$ given $X$ is

$$\hat{Y} = \frac{1}{\mu_3^2 - \mu_4 + 1}(aX^2 + bX + c)$$

where

$$\begin{aligned} a &= \mu_3\beta_1 - \beta_2, \\ b &= (1 - \mu_4)\beta_1 + \mu_3\beta_2, \\ c &= -\mu_3\beta_1 + \beta_2. \end{aligned}$$

Your task is to prove the Projection Property for $\hat{Y}$.

(a) Prove that $\mathbb{E}[Y - \hat{Y}] = 0$.

(b) Prove that $\mathbb{E}[(Y - \hat{Y})X] = 0$.

(c) Prove that $\mathbb{E}[(Y - \hat{Y})X^2] = 0$.

Any quadratic function of $X$ is a linear combination of 1, $X$, and $X^2$. Hence, these equations together imply that $Y - \hat{Y}$ is orthogonal to any quadratic function of $X$, and so $\hat{Y}$ is the best quadratic estimator of $Y$.

**Solution:**

(a) By linearity of expectation:

$$\mathbb{E}[Y - \hat{Y}] = \mathbb{E}[Y] - \frac{a\,\mathbb{E}[X^2] + b\,\mathbb{E}[X] + c}{\mu_3^2 - \mu_4 + 1} = \frac{-a - c}{\mu_3^2 - \mu_4 + 1} = 0$$

since $\mathbb{E}[X] = \mathbb{E}[Y] = 0$ and $\mathbb{E}[X^2] = 1$.

(b)

$$\begin{aligned}\mathbb{E}[(Y - \hat{Y})X] &= \mathbb{E}[XY] - \frac{a\,\mathbb{E}[X^3] + b\,\mathbb{E}[X^2] + c\,\mathbb{E}[X]}{\mu_3^2 - \mu_4 + 1} \\ &= \beta_1 - \frac{(\mu_3\beta_1 - \beta_2)\mu_3 + ((1-\mu_4)\beta_1 + \mu_3\beta_2)}{\mu_3^2 - \mu_4 + 1}\end{aligned}$$

which, after a little algebra, gives 0.

(c)

$$\begin{aligned}\mathbb{E}[(Y - \hat{Y})X^2] &= \mathbb{E}[X^2 Y] - \frac{a\,\mathbb{E}[X^4] + b\,\mathbb{E}[X^3] + c\,\mathbb{E}[X^2]}{\mu_3^2 - \mu_4 + 1} \\ &= \beta_2 - \frac{\mu_4(\mu_3\beta_1 - \beta_2) + \mu_3((1-\mu_4)\beta_1 + \mu_3\beta_2) - \mu_3\beta_1 + \beta_2}{\mu_3^2 - \mu_4 + 1}\end{aligned}$$

which, after a little algebra, gives 0.

# 4 Marbles in a Bag

We have $r$ red marbles, $b$ blue marbles, and $g$ green marbles in the same bag. If we sample marbles with replacement until we get 3 red marbles (not necessarily consecutively), how many blue marbles should we expect to see? (*Hint*: It might be useful to use Law of Total Expectation, $E(Y) = E(E(Y|X))$ )

**Solution:**

Let $Y$ be the number of blue marbles we see. Let $X$ be the samples we take until we get 3 red marbles.

Let us first compute $\mathbb{E}[Y \mid X]$. Let $Y_i$ be 1 if we see a blue marble on the $i$th sample and $Y = \sum_{i=1}^{k} Y_i$.

This means

$$\begin{aligned}\mathbb{E}[Y \mid X=k] &= \mathbb{E}\left[\sum_{i=1}^{k} Y_i \,\middle|\, X=k\right] \\ &= \sum_{i=1}^{k} \mathbb{E}[Y_i \mid X=k].\end{aligned}$$

However, three $Y_i$ (call them $Y_a, Y_b, Y_c$) have $\mathbb{E}[Y_i]=0$, since there are necessarily 3 red marbles. This means the other $k-3$ marbles are necessarily blue or green.

$$\begin{aligned}\sum_{i\neq a,b,c} \mathbb{E}[Y_i \mid X=k] &= \sum_{i\neq a,b,c} \mathbb{P}[Y_i=1 \mid X=k] \\ &= \sum_{i\neq a,b,c} \frac{b}{b+g} \\ &= (k-3)\frac{b}{b+g}.\end{aligned}$$

This means

$$\mathbb{E}[Y \mid X] = (X-3)\frac{b}{b+g}.$$

Using the Law of Total Expectation, we know that

$$\begin{aligned}\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y \mid X]] = \mathbb{E}\left[\frac{b}{b+g}(X-3)\right] &= \frac{b}{b+g}\,\mathbb{E}[X-3] \\ &= \frac{b}{b+g}(\mathbb{E}[X]-3).\end{aligned}$$

We notice that

$$X = X_1 + X_2 + X_3,$$

where each $X_i$ represents the number of marbles seen between drawing the $(i-1)$th and $i$th red marble. We know that the absolute number of marbles seen between 2 consecutive red marbles is geometric, since we want to find the number of draws until the first red marble.

$$X_i \sim \text{Geometric}\left(\frac{r}{r+b+g}\right)$$

Since $X_1, X_2, X_3$ are identically distributed, we know that

$$\mathbb{E}[X] = 3\,\mathbb{E}[X_i] = 3(r+g+b)/r$$

.

$$\mathbb{E}[Y] = \frac{b}{b+g}\left(3\frac{r+g+b}{r} - 3\right) = \frac{3b}{r}.$$

**Alternate Solution**:

We know that the absolute number of marbles $N$ seen between 2 consecutive red marbles is geometric, since we want to find the number of draws until the first red marble. And given the number of marbles, $N$, between 2 consecutive red marbles, the number of blue marbles among these is distributed binomially.

Therefore, each $X_i$ is drawn from a binomial distribution, where the number of trials is distributed geometrically.

We exclude the very last marble in the binomial distribution, because we know it must be red (and therefore cannot be blue). And the probability for the binomial is $b/(b+g)$ because we know that in between 2 consecutive red balls, we can only have blue or green balls. So,

$$X_i \sim \text{Binomial}\left(\frac{b}{b+g}, N-1\right), \qquad \text{where} \qquad N \sim \text{Geometric}\left(\frac{r}{r+b+g}\right).$$

And, applying the law of conditional expectation, we have

$$\begin{aligned}
\mathbb{E}(X_i) &= \mathbb{E}(\mathbb{E}(X_i \mid N)) \\
&= \mathbb{E}\left((N-1)\frac{b}{b+g}\right) \\
&= \frac{b}{b+g}\mathbb{E}(N-1) \\
&= \frac{b}{b+g}\left(\frac{r+b+g}{r} - 1\right) \\
&= \frac{b}{b+g}\left(\frac{b+g}{r}\right) \\
&= \frac{b}{r}.
\end{aligned}$$

We know that each of the $X_i$'s is identically distributed, so

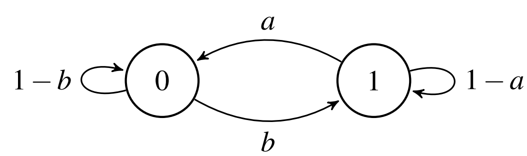
$$\mathbb{E}(X) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \mathbb{E}(X_3) = 3 \cdot \mathbb{E}(X_1) = \frac{3b}{r}.$$

# 5 Markov Chain Terminology

In this question, we will walk you through terms related to Markov chains.

1. (Irreducibility) A Markov chain is irreducible if, starting from any state $i$, the chain can transition to any other state $j$, possibly in multiple steps.

2. (Periodicity) $d(i) := \gcd\{n > 0 \mid P^n(i,i) = \mathbb{P}[X_n = i \mid X_0 = i] > 0\}$, $i \in \mathscr{X}$. If $d(i) = 1 \; \forall i \in \mathscr{X}$, then the Markov chain is aperiodic; otherwise it is periodic.
3. (Matrix Representation) Define the transition probability matrix $P$ by filling entry $(i, j)$ with probability $P(i, j)$.
4. (Invariance) A distribution $\pi$ is invariant for the transition probability matrix $P$ if it satisfies the following balance equations: $\pi = \pi P$.




(a) For what values of $a$ and $b$ is the above Markov chain irreducible? Reducible?

(b) For $a = 1$, $b = 1$, prove that the above Markov chain is periodic.

(c) For $0 < a < 1$, $0 < b < 1$, prove that the above Markov chain is aperiodic.

(d) Construct a transition probability matrix using the above Markov chain.

(e) Write down the balance equations for this Markov chain and solve them. Assume that the Markov chain is irreducible.

**Solution:**

(a) The Markov chain is irreducible if both $a$ and $b$ are non-zero. It is reducible if at least one is 0.

(b) We compute $d(0)$ to find that:

$$d(0) = \gcd\{2, 4, 6, \ldots\} = 2.$$

Thus, the chain is periodic.

(c) We compute $d(0)$ to find that:

$$d(0) = \gcd\{1, 2, 3, \ldots\} = 1.$$

Thus, the chain is aperiodic.

(d)

$$\begin{bmatrix} 1-b & b \\ a & 1-a \end{bmatrix}$$

(e)

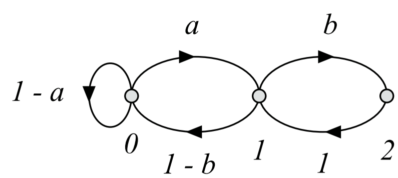
$$\pi(0) = (1-b)\pi(0) + a\pi(1),$$
$$\pi(1) = b\pi(0) + (1-a)\pi(1).$$

One of the equations is redundant. We throw out the second equation and replace it with $\pi(0)+\pi(1) = 1$. This gives the solution

$$\pi = \frac{1}{a+b}\begin{bmatrix} a & b \end{bmatrix}.$$

# 6 Analyze a Markov Chain

Consider the Markov chain $X(n)$ with the state diagram shown below where $a, b \in (0,1)$.




(a) Show that this Markov chain is aperiodic;

(b) Calculate $\mathbb{P}[X(1) = 1, X(2) = 0, X(3) = 0, X(4) = 1 \mid X(0) = 0]$;

(c) Calculate the invariant distribution;

(d) Let $T_i = \min\{n \geq 0 \mid X(n) = i\}$, $T_i$ is the number of steps until we transit to state $i$ for the first time. Calculate $\mathbb{E}[T_2 \mid X(0) = 1]$.

**Solution:**

(a) The Markov chain is irreducible because $a, b \in (0,1)$. Also, $P(0,0) > 0$, so that

$$\gcd\{n > 0 \mid P^n(0,0) > 0\} = \gcd\{1,2,3,\ldots\} = 1,$$

which shows that the Markov chain is aperiodic.
We can also notice from the definition of aperiodic Markov chain, that if a Markov chain has a self loop with nonzero probability ($P(0,0) > 0$ in this example), which says the smallest number of steps from a state to itself if 1, it is aperiodic.

(b) As a result of the Markov property, we know our state at timestep $n$ depends only on timestep $n-1$. We see that the probability is

$$P(0,1) \times P(1,0) \times P(0,0) \times P(0,1) = a(1-b)(1-a)a.$$

(c) The balance equations are

$$\pi(0) = (1-a)\pi(0) + (1-b)\pi(1)$$
$$\pi(1) = a\pi(0) + \pi(2).$$

After some simple manipulations, we see that they imply the following equations:

$$a\pi(0) = (1-b)\pi(1)$$
$$b\pi(1) = \pi(2).$$

These equations express the equality of the probability of a jump from $i$ to $i+1$ and from $i+1$ to $i$, for $i=0$ and $i=1$, respectively. These relations are called the "detailed balance equations". From these equations we find successively that

$$\pi(1) = \frac{a}{1-b}\pi(0) \text{ and } \pi(2) = b\pi(1) = \frac{ab}{1-b}\pi(0).$$

The normalization equation is

$$1 = \pi(0) + \pi(1) + \pi(2) = \pi(0)\left[1 + \frac{a}{1-b} + \frac{ab}{1-b}\right]$$
$$= \pi(0)\frac{1-b+a+ab}{1-b},$$

so that

$$\pi(0) = \frac{1-b}{1-b+a+ab}.$$

Thus,

$$\pi = \frac{1}{1-b+a+ab}\begin{bmatrix}1-b & a & ab\end{bmatrix}.$$

(d) We define

$$\beta(i) = \mathbb{E}[T_2 \mid X(0) = i], i = 0,1,2.$$

The FSE are $\beta(2) = 0$ and

$$\beta(0) = 1 + (1-a)\beta(0) + a\beta(1)$$
$$\beta(1) = 1 + (1-b)\beta(0).$$

The first equation is equivalent to

$$\beta(0) = \frac{1}{a} + \beta(1).$$

Substituting this expression in the second equation, we get

$$\beta(1) = 1 + (1-b)\left(\frac{1}{a} + \beta(1)\right) = (1-b)\beta(1) + \frac{1+a-b}{a},$$

so that

$$\beta(1) = \frac{1+a-b}{ab}.$$