1 Continuous LLSE

Suppose that X and Y are uniformly distributed on the shaded region in the figure below.

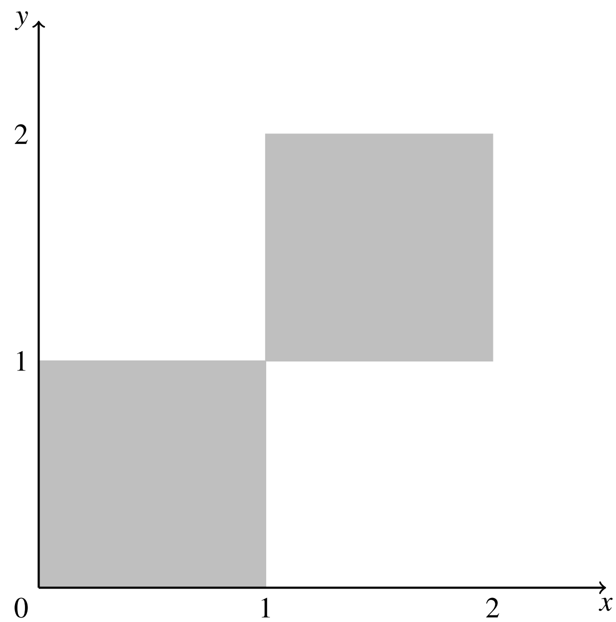


Figure 1: The joint density of (X,Y) is uniform over the shaded region.

That is, X and Y have the joint distribution:

$$f_{X,Y}(x,y) = \begin{cases} 1/2, & 0 \le x \le 1, 0 \le y \le 1 \\ 1/2, & 1 \le x \le 2, 1 \le y \le 2 \end{cases}$$

(a) Do you expect X and Y to be positively correlated, negatively correlated, or neither?

(b) Compute the marginal distribution of X.

(c) Compute $L[Y \mid X]$.

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

Solution:

(a) Positively correlated, because high values of Y correspond to high values of X.

(b) Intuitively, if we slice the joint distribution at any $x \in [0,2]$, then the probability is the same, so we should expect X to be uniformly distributed on $[0,2]$. We verify this by explicit computation:

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,\mathrm{d}y = 1\{0 \le x \le 1\} \int_0^1 \frac{1}{2}\,\mathrm{d}y + 1\{1 \le x \le 2\} \int_1^2 \frac{1}{2}\,\mathrm{d}y \\ &= \frac{1}{2} 1\{0 \le x \le 2\} \end{aligned}$$

(c) $\mathbb{E}[X] = \mathbb{E}[Y] = 1$ by symmetry. Since X is uniform on $[0,2]$, $\operatorname{var}(X) = 4 \cdot 1/12 = 1/3$ (since the variance of a $U[0,1]$ random variable is $1/12$). We compute the covariance:

$$\begin{aligned} \mathbb{E}[XY] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int_0^1 \int_0^1 xy \cdot \frac{1}{2}\,\mathrm{d}x\,\mathrm{d}y + \int_1^2 \int_1^2 xy \cdot \frac{1}{2}\,\mathrm{d}x\,\mathrm{d}y \\ &= \frac{1}{2} \left(\int_0^1 x\,\mathrm{d}x \int_0^1 y\,\mathrm{d}y + \int_1^2 x\,\mathrm{d}x \int_1^2 y\,\mathrm{d}y \right) = \frac{1}{2} \left(\frac{1}{4} + \frac{9}{4} \right) = \frac{5}{4} \end{aligned}$$

So $\operatorname{cov}(X,Y) = 5/4 - 1 \cdot 1 = 1/4$. The LLSE is

$$\begin{aligned} L[Y \mid X] - 1 &= \frac{1/4}{1/3}(X - 1) \\ L[Y \mid X] &= \frac{3}{4} X + \frac{1}{4} \end{aligned}$$

(d) The easiest way to solve this is to look at the picture of the joint density, and draw horizontal lines through middles of each of the two squares. Intuitively, $\mathbb{E}[Y \mid X]$ means "for each slice of $X = x$, what is the best guess of Y"? Slightly more formally, one can argue that conditioned on $X = x$ for $0 < x < 1$, $Y \sim U[0,1]$, so $\mathbb{E}[Y \mid X = x] = 1/2$ in this region. Conditioned on $X = x$ for $1 < x < 2$, $Y \sim U[1,2]$, so $\mathbb{E}[Y \mid X = x] = 3/2$ in this region. See Figure 2.

$$\mathbb{E}[Y \mid X = x] = \begin{cases} 1/2, & 0 \le x \le 1 \\ 3/2, & 1 \le x \le 2 \end{cases}$$

2 Markov Chain Basics

A Markov chain is a sequence of random variables X_n, $n = 0, 1, 2, \dots$. Here is one interpretation of a Markov chain: X_n is the state of a particle at time n. At each time step, the particle can jump to another state. Formally, a Markov chain satisfies the Markov property:

$$\mathbb{P}(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = \mathbb{P}(X_{n+1} = j \mid X_n = i), \tag{1}$$

for all n, and for all sequences of states $i_0, \dots, i_{n-1}, i, j$. In other words, the Markov chain does not have any memory; the transition probability only depends on the current state, and not the history of states that have been visited in the past.

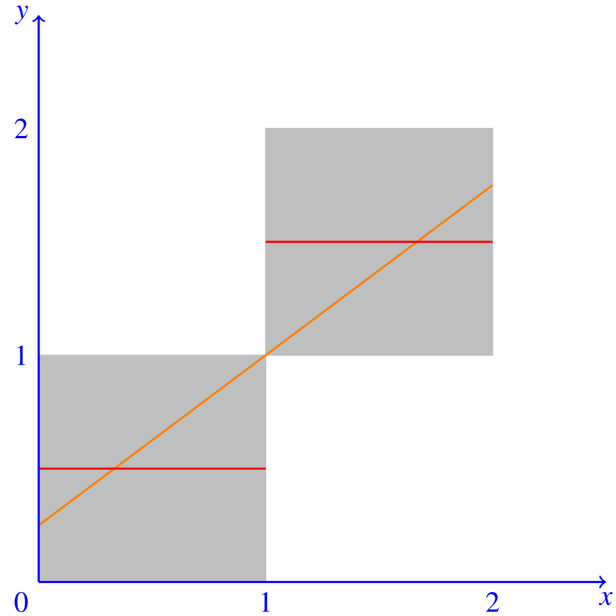


Figure 2: The LLSE is the orange line. The MMSE is the red function.

(a) In lecture, we learned that we can specify Markov chains by providing three ingredients: $\mathscr{X}$, P, and π_0. What do these represent, and what properties must they satisfy?

(b) If we specify $\mathscr{X}$, P, and π_0, we are implicitly defining a sequence of random variables X_n, $n = 0, 1, 2, \ldots$, that satisfies (1). Explain why this is true.

(c) Calculate $\mathbb{P}(X_1 = j)$ in terms of π_0 and P. Then, express your answer in matrix notation. What is the formula for $\mathbb{P}(X_n = j)$ in matrix form?

Solution:

(a) $\mathscr{X}$ is the set of states, which is the range of possible values for X_n. In this course, we only consider finite $\mathscr{X}$.

P contains the transition probabilities. $P(i, j)$ is the probability of transitioning from state i to state j. It must satisfy $\sum_{j \in \mathscr{X}} P(i, j) = 1 \; \forall i \in \mathscr{X}$, which says that the probability that *some* transition occurs must be 1. Also, the entries must be non-negative: $P(i, j) \geq 0 \; \forall i, j \in \mathscr{X}$. A matrix satisfying these two properties is called a stochastic matrix.

Note that we allow states to transition to themselves, i.e. it is possible for $P(i, i) > 0$.

π_0 is the initial distribution, that is, $\pi_0(i) = \mathbb{P}(X_0 = i)$. Similarly, we let π_n be the distribution of X_n. Since π_0 is a probability distribution, its entries must be non-negative and $\sum_{i \in \mathscr{X}} \pi_0(i) = 1$.

(b) The sequence of random variables X_n, $n = 0, 1, 2, \ldots$, is defined in the following way:

- X_0 has distribution π_0, i.e. $\mathbb{P}(X_0 = i) = \pi_0(i)$.
- X_{n+1} has distribution given by

$$\mathbb{P}(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = \mathbb{P}(X_{n+1} = j \mid X_n = i) = P(i, j),$$

for all $n = 0, 1, 2, \ldots$.

It is important to realize the connection between the Markov property (1) and the transition matrix P. P contains information about the transition probabilities in one step. If the Markov property did not hold, then P would not be enough to specify the distribution of X_{n+1}. Conversely, if we only specify P, then we are implicitly assuming that the transition probabilities do not depend on anything other than the current state.

(c) By the Law of Total Probability,

$$\mathbb{P}(X_1 = j) = \sum_{i \in \mathscr{X}} \mathbb{P}(X_1 = j, X_0 = i) = \sum_{i \in \mathscr{X}} \mathbb{P}(X_0 = i)\mathbb{P}(X_1 = j \mid X_0 = i) = \sum_{i \in \mathscr{X}} \pi_0(i)P(i, j).$$

If we write $\pi_1(j) = \mathbb{P}(X_1 = j)$ and π_0 as row vectors, then in matrix notation we have

$$\pi_1 = \pi_0 P.$$

The effect of a transition is right-multiplication by P. After n time steps, we have

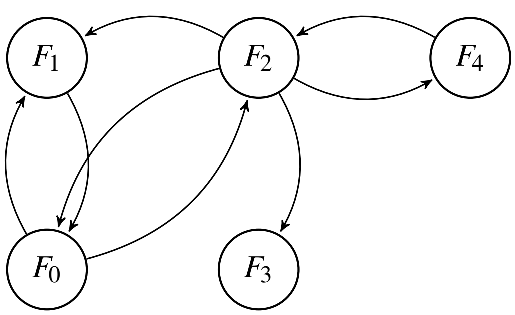
$$\pi_n = \pi_0 P^n.$$

At this point, it should be mentioned that many calculations involving Markov chains are very naturally expressed with the language of matrices. Consequently, Markov chains are very well-suited for computers, which is one of the reasons why Markov chain models are so popular in practice.

3 The Dwinelle Labyrinth

You have decided to take a humanities class this semester, a French class to be specific. Instead of a final exam, your professor has issued a final paper. You must turn in this paper *before* noon to the professor's office on floor 3 in Dwinelle, and it's currently 11:48 a.m.

Let Dwinelle be modeled by the following Markov chain. Instead of rushing to turn it in, we will spend valuable time computing whether or not we *could have* made it. Suppose walking between floors takes 1 minute.



(a) Will you make it in time if you choose a floor to transition to uniformly at random? (If T_i is the number of steps needed to get to F_3 starting from F_i, where $i \in \{0, 1, 2, 3, 4\}$, is $\mathbb{E}[T_0] < 12$?)

(b) Will you make it in time, if for every floor, you order all accessible floors and are twice as likely to take higher floors? (If you are considering 1, 2, or 3, you will take each with probabilities $1/7$, $2/7$, $4/7$, respectively.)

Solution:

(a) Write out all of the first-step equations.

$$\begin{aligned}
\mathbb{E}[T_0] &= 1 + \frac{1}{2}\mathbb{E}[T_1] + \frac{1}{2}\mathbb{E}[T_2] \\
\mathbb{E}[T_1] &= 1 + \mathbb{E}[T_0] \\
\mathbb{E}[T_2] &= 1 + \frac{1}{4}\mathbb{E}[T_0] + \frac{1}{4}\mathbb{E}[T_1] + \frac{1}{4}\mathbb{E}[T_3] + \frac{1}{4}\mathbb{E}[T_4] \\
\mathbb{E}[T_3] &= 0 \\
\mathbb{E}[T_4] &= 1 + \mathbb{E}[T_2]
\end{aligned}$$

Let us rewrite these equations, before placing it in matrix form.

$$\begin{aligned}
-1 &= -\mathbb{E}[T_0] + \frac{1}{2}\mathbb{E}[T_1] + \frac{1}{2}\mathbb{E}[T_2] \\
-1 &= -\mathbb{E}[T_1] + \mathbb{E}[T_0] \\
-1 &= -\mathbb{E}[T_2] + \frac{1}{4}\mathbb{E}[T_0] + \frac{1}{4}\mathbb{E}[T_1] + \frac{1}{4}\mathbb{E}[T_3] + \frac{1}{4}\mathbb{E}[T_4] \\
0 &= \mathbb{E}[T_3] \\
-1 &= -\mathbb{E}[T_4] + \mathbb{E}[T_2]
\end{aligned}$$

We can rewrite this in matrix form.

$$P = \begin{bmatrix} -1 & 1/2 & 1/2 & 0 & 0 & -1 \\ 1 & -1 & 0 & 0 & 0 & -1 \\ 1/4 & 1/4 & -1 & 1/4 & 1/4 & -1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & -1 \end{bmatrix}$$

We can now reduce the matrix.

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 15 \\ 0 & 1 & 0 & 0 & 0 & 16 \\ 0 & 0 & 1 & 0 & 0 & 12 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 13 \end{bmatrix}$$

We see that $\mathbb{E}[T_0] = 15$, meaning it will take 15 minutes for us to get to floor 3. Unfortunately, we only have 12 minutes.

(b) Write out all of the first-step equations.

$$\begin{aligned}
\mathbb{E}[T_0] &= 1 + \frac{1}{3}\mathbb{E}[T_1] + \frac{2}{3}\mathbb{E}[T_2] \\
\mathbb{E}[T_1] &= 1 + \mathbb{E}[T_0] \\
\mathbb{E}[T_2] &= 1 + \frac{1}{15}\mathbb{E}[T_0] + \frac{2}{15}\mathbb{E}[T_1] + \frac{4}{15}\mathbb{E}[T_3] + \frac{8}{15}\mathbb{E}[T_4] \\
\mathbb{E}[T_3] &= 0 \\
\mathbb{E}[T_4] &= 1 + \mathbb{E}[T_2]
\end{aligned}$$

Let us rewrite these equations, before placing it in matrix form.

$$\begin{aligned}
-1 &= -\mathbb{E}[T_0] + \frac{1}{3}\mathbb{E}[T_1] + \frac{2}{3}\mathbb{E}[T_2] \\
-1 &= -\mathbb{E}[T_1] + \mathbb{E}[T_0] \\
-1 &= -\mathbb{E}[T_2] + \frac{1}{15}\mathbb{E}[T_0] + \frac{2}{15}\mathbb{E}[T_1] + \frac{4}{15}\mathbb{E}[T_3] + \frac{8}{15}\mathbb{E}[T_4] \\
0 &= \mathbb{E}[T_3] \\
-1 &= -\mathbb{E}[T_4] + \mathbb{E}[T_2]
\end{aligned}$$

We can rewrite this in matrix form.

$$P = \begin{bmatrix} -1 & 1/3 & 2/3 & 0 & 0 & -1 \\ 1 & -1 & 0 & 0 & 0 & -1 \\ 1/15 & 2/15 & -1 & 4/15 & 8/15 & -1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & -1 \end{bmatrix}$$

We row reduce to get the following.

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 9.75 \\ 0 & 1 & 0 & 0 & 0 & 10.75 \\ 0 & 0 & 1 & 0 & 0 & 7.75 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 8.75 \end{bmatrix}$$

We see that $\mathbb{E}[T_0] = 9.75$, meaning it will take 9.75 minutes for us to get to floor 3. That's fewer than 12 minutes, so if you finished this computation in less than 2 minutes and 15 seconds, you could make it!