Math 456: Mathematical Modeling

Thursday, February 15th 2018

Markov Chains,
and the Chapman-Kolmogorov equation

Thursday, February 15th 2018

Gambler's Ruin

Sum of independent Bernoulli Random Variables

Let $Y_1, Y_2, Y_3, \ldots$ be all **independent** random variables with

$$\mathrm{P}(Y_n = 1) = 0.6 \quad \mathrm{P}(Y_n = -1) = 0.4 \text{ for each } n$$

Fix $x \in \mathbb{N}$ and define for each n,

$$X_n = x + \sum_{k=1}^{n} Y_k$$

Interpretation: you are playing a game where on each step a coin is flipped and you either win a dollar $Y_n = 1$ or lose a dollar $Y_n = -1$, x is your initial dollar amount, and X_n is the number of dollars you have after n steps.

Gambler's Ruin

Sum of independent Bernoulli Random Variables

The random variables $X_1, X_2, X_3, \ldots$ are no longer independent, however, they have an important property...

Consider the probability of X_{n+1} being of a given value provided X_n is known

$$\mathrm{P}(X_{n+1} = j \mid X_n = i)$$

Let's compute this!

Gambler's Ruin

Sum of independent Bernoulli Random Variables

Since $X_{n+1} = X_n + Y_{n+1}$ by definition,

$$\mathrm{P}(X_{n+1} = j \mid X_n = i) = \mathrm{P}(X_n + Y_{n+1} = j \mid X_n = i)$$

Gambler's Ruin

Sum of independent Bernoulli Random Variables

Since $X_{n+1} = X_n + Y_{n+1}$ by definition,

$$\mathrm{P}(X_{n+1} = j \mid X_n = i) = \mathrm{P}(X_n + Y_{n+1} = j \mid X_n = i)$$

$$\begin{aligned} \Rightarrow \mathrm{P}(X_{n+1} = j \mid X_n = i) &= \mathrm{P}(i + Y_{n+1} = j \mid X_n = i) \\ &= \mathrm{P}(Y_{n+1} = j - i \mid X_n = i) \end{aligned}$$

Gambler's Ruin

Sum of independent Bernoulli Random Variables

Since $X_{n+1} = X_n + Y_{n+1}$ by definition,

$$\mathrm{P}(X_{n+1} = j \mid X_n = i) = \mathrm{P}(X_n + Y_{n+1} = j \mid X_n = i)$$

$$\begin{aligned} \Rightarrow \mathrm{P}(X_{n+1} = j \mid X_n = i) &= \mathrm{P}(i + Y_{n+1} = j \mid X_n = i) \\ &= \mathrm{P}(Y_{n+1} = j - i \mid X_n = i) \end{aligned}$$

Since Y_{n+1} is independent of $Y_1, \ldots, Y_{n-1}$, and Y_n, we have

$$\mathrm{P}(X_{n+1} = j \mid X_n = i) = \mathrm{P}(Y_{n+1} = j - i)$$

Gambler's Ruin

Sum of independent Bernoulli Random Variables

In conclusion,

$$\mathrm{P}(X_{n+1} = i + 1 \mid X_n = i) = 0.6$$
$$\mathrm{P}(X_{n+1} = i - 1 \mid X_n = i) = 0.4$$

Moreover,

$$\mathrm{P}(X_{n+1} = j \mid X_n = i) = 0, \text{ if } j \neq i \pm 1$$

THEREFORE:

knowledge of X_n **simplifies** the distribution of X_{n+1}.

Gambler's Ruin

knowledge of X_n **simplifies** the distribution of X_{n+1}.

Do we gain anything if besides knowing the value of X_n, we also know some of the previous values X_{n-1}, $X_{n-2}, \ldots$?

Gambler's Ruin

knowledge of X_n **simplifies** the distribution of X_{n+1}.

Do we gain anything if besides knowing the value of X_n, we also know some of the previous values X_{n-1}, $X_{n-2}, \ldots$?

Answer: No!.

Gambler's Ruin

In fact: given steps $k_1, \ldots, k_m$ all previous to n, and any values $\alpha_1, \ldots, \alpha_m$, we have that

$$\mathrm{P}(X_{n+1} = j \mid X_n = i, X_{k_1} = \alpha_1, \ldots, X_{k_m} = \alpha_m)$$

is always equal to

$$\mathrm{P}(X_{n+1} = j \mid X_n = i).$$

Gambler's Ruin

In fact: given steps $k_1, \ldots, k_m$ all previous to n, and any values $\alpha_1, \ldots, \alpha_m$, we have that

$$\mathrm{P}(X_{n+1} = j \mid X_n = i, X_{k_1} = \alpha_1, \ldots, X_{k_m} = \alpha_m)$$

is always equal to

$$\mathrm{P}(X_{n+1} = j \mid X_n = i).$$

To put it in rather poetic terms:

The future of the system is determined entirely by it's present, regardless of the past history that led to its present state.

Gambler's Ruin

Let us make **one modification** and propose a **game**: you are given some target money $M \in \mathbb{N}$ you want to reach before leaving the game. You keep playing that game until you reach M or until you run out of money, this happens at the time step T defined by

$$T = \min\{n \mid X_n \in \{0, M\}\}$$

at which point you stop the game and stay with X_T dollars (so, $X_T = X_{T+1} = X_{T+2}$ and so on).

Question: How much money would you be willing to pay to play the game, based on M and the initial capital x?

Gambler's Ruin

So let us still denote by X_n this new modified game. X_n is no longer a sum of independendent Bernoulli random variables. But we still have this property:

Gambler's Ruin

So let us still denote by X_n this new modified game. X_n is no longer a sum of independendent Bernoulli random variables. But we still have this property:

Given steps $k_1, \ldots, k_m$ all previous to n, and any values $\alpha_1, \ldots, \alpha_m$, we have that

$$\mathrm{P}(X_{n+1} = j \mid X_n = i, X_{k_1} = \alpha_1, \ldots, X_{k_m} = \alpha_m)$$

is always equal to

$$\mathrm{P}(X_{n+1} = j \mid X_n = i).$$

Gambler's Ruin

The difference is now we have: $X_n \in \{0, 1 \ldots, M\}$ for all n, with

$$
\begin{aligned}
P(X_{n+1} = 0 \mid X_n = 0) = 1 \\
P(X_{n+1} = M \mid X_n = M) = 1
\end{aligned}
$$

and $P(X_{n+1} = j \mid X_n = i)$ being as before when $0 < i < M$.

The Markov Propery

The **Markov property** is a condition on a sequence of random variables $X_1, X_2, \ldots$ thought to represent the state of a system over time (n representing time).

It means to capture a feature described earlier as

The future of the system is determined entirely by it's present, regardless of the past history that led to its present state.

The Markov Property

A sequence $X_1, X_2, X_3, \ldots$ is said to be a **Markov Chain** if

$$\mathrm{P}(X_{n+1} = j \mid X_n = i, X_{k_1} = \alpha_1, \ldots, X_{k_m} = \alpha_m)$$

(*where the indices* $k_1, \ldots, k_m$ *are all strictly less than* n)

is a number entirely determined by n, i, and j. This number is called **the transition probability** from i to j at time n.

If the number is also independent of n, then we say the Markov chain is **homogeneous** and **the transition probability** is denoted simply $\mathrm{p}(i, j)$.

The Markov Property

Transition Probability

Let us think of the states as being labeled from 1 through N.
Then, the transition probability requires $N \times N$ numbers $\mathrm{p}(i, j)$

Given any i, we have

$$\sum_j \mathrm{p}(i, j) = 1$$

Question: Why is this so?

The Markov Property

Transition Probability

The total probability formula says that

$$\begin{aligned} \mathrm{P}(X_{n+1} = j) &= \sum_k \mathrm{P}(X_{n+1} = j \mid X_n = k)\mathrm{P}(X_n = k) \\ &= \sum_k \mathrm{p}(k, j)\mathrm{P}(X_n = k) \end{aligned}$$

This has a linear algebra interpretation: if we introduce a transition probability matrix with entries $\mathrm{p} = (\mathrm{p}(i, j))_{ij}$, then to go from the distribution of X_n to the distribution of X_{n+1}, all we need to do is multiply the "vector" of probabilities by the **adjoint** of the matrix p.

The Markov Property

Transition Probability

*...all we need to do is multiply the "vector" of probabilities by the **adjoint** of the matrix* p*.*

This observation leads to what is known as the **Chapman-Kolmogorov** equation, which allows us to compute probabilities of the state several time steps into the future.

The Markov Property

One more time

The future of the system is determined entirely by it's present, regardless of the past history that led to its present state.

The Markov Property

One more time

Let us also repeat the definition of a **Markov Chain**.

Take a sequence $X_1, X_2, X_3, \ldots$ of random variables.
This sequence is said to be a **Markov Chain** if

$$\mathrm{P}(X_{n+1} = j \mid X_n = i, X_{k_1} = \alpha_1, \ldots, X_{k_m} = \alpha_m)$$

(*where the indices* $k_1, \ldots, k_m$ *are all strictly less than* n)

is a number entirely determined by n, i, and j. This number is called **the transition probability** from i to j at time n.

The Markov Property

Transition probability matrix

If the number is also independent of n, then we say the Markov chain is **homogeneous** and **the transition probability** is denoted simply $\mathrm{p}(i, j)$.

The i, j *and the* α_k*'s here denote possible states of the system, they could be numbers, or any other kind of label for the states.*

For convenience, let us think of the states as being labeled from 1 through N, so that $i, j = 1, \ldots, N$.

We may think of the transition probabilities $\mathrm{p}(i, j)$ as arranged along rows and columns corresponding to i and j.

Examples of Markov Chains

Example 1: (Gambler's Ruin with $M = 5$)

This is a system with 6 possible states, labeled 0 through 5. If the coin is fair, the probabilities are

	0	**1**	**2**	**3**	**4**	**5**
0	1	0	0	0	0	0
1	0.5	0	0.5	0	0	0
2	0	0.5	0	0.5	0	0
3	0	0	0.5	0	0.5	0
4	0	0	0	0.5	0	0.5
5	0	0	0	0	0	1

Examples of Markov Chains

Example 1: (Gambler's Ruin with $M = 5$)

This arrangement is no accident.
It's natural –and extremely useful– to think of the transition probabilities as forming a $N \times N$ matrix

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

This is called the **transition matrix** for the chain.

Examples of Markov Chains

Example 1: (Gambler's Ruin with $M = 5$)

All the **essential information about the Markov chain** is contained in the transition matrix.

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Examples of Markov Chains

Example 1: (Gambler's Ruin with $M = 5$)

If the coin were biased, we would have two numbers $p, q \in [0, 1]$ with $p + q = 1$ $p \neq q$, and the transition matrix is

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ p & 0 & q & 0 & 0 & 0 \\ 0 & p & 0 & q & 0 & 0 \\ 0 & 0 & p & 0 & q & 0 \\ 0 & 0 & 0 & p & 0 & q \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Examples of Markov Chains

Example 2: (a two-state model for weather)

Denote 1 = **Rainy Day**, 2 = **Sunny Day**.

	1	**2**
1	0.6	0.4
2	0.2	0.8

For Rainy days:
Next day will be Rainy or Sunny with probalities 0.6 and 0.4.
For Sunny days:
Next day will be Rainy or Sunny with probalities 0.2 and 0.8.

Examples of Markov Chains

Example 2: (a two-state model for weather)

Again, we arrange these probabilities in a matrix,

$$\begin{pmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{pmatrix}$$

Since the system only has two states, the matrix is 2×2.

Question: Out of N days, what proportion of them are rainy?

Examples of Markov Chains

Example 2: (a two-state model for weather)

$$\begin{pmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{pmatrix}$$

Fix i and j, what is the probability of going from i to j in two steps? i.e. what is $\mathrm{P}(X_2 = j \mid X_0 = i)$?

Examples of Markov Chains

Example 2: (a two-state model for weather)

$$\begin{pmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{pmatrix}$$

Well, there are two ways to go from i to j in exactly **two** steps,

$$i \mapsto 1 \mapsto j$$
$$i \mapsto 2 \mapsto j$$

Then, $\mathrm{P}(X_2 = j \mid X_0 = i)$ is equal to

$$\mathrm{P}(X_2 = j, X_1 = 1 \mid X_0 = i) + \mathrm{P}(X_2 = j, X_1 = 2 \mid X_0 = i)$$

Examples of Markov Chains

Example 2: (a two-state model for weather)

$$\begin{pmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{pmatrix}$$

In terms of conditional probabilities, this sum is equal to

$$P(X_2 = j \mid X_0 = i, X_1 = 1)\mathrm{P}(X_1 = 1 \mid X_0 = i)$$
$$+ \mathrm{P}(X_2 = j \mid X_0 = i, X_1 = 2)\mathrm{P}(X_1 = 2 \mid X_0 = i)$$

Then, the Markov property says this is equal to

$$\mathrm{P}(X_2 = j \mid X_1 = 1)\mathrm{P}(X_1 = 1 \mid X_0 = i)$$
$$+ \mathrm{P}(X_2 = j \mid X_1 = 2)\mathrm{P}(X_1 = 2 \mid X_0 = i)$$

Examples of Markov Chains

Example 2: (a two-state model for weather)

$$\begin{pmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{pmatrix}$$

In other words, $\mathrm{P}(X_2 = j \mid X_0 = i)$ is given by

$$\mathrm{p}(1, j)\mathrm{p}(i, i) + \mathrm{p}(2, j)\mathrm{p}(i, 2)$$

Examples of Markov Chains

Example 2: (a two-state model for weather)

$$\begin{pmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{pmatrix}$$

Let us write $\mathrm{p}^2(i,j) = \mathrm{P}(X_{n+2} = j \mid X_2 = i)$, then

$$\mathrm{p}^2(i,j) = \mathrm{p}(1,j)\mathrm{p}(i,i) + \mathrm{p}(2,j)\mathrm{p}(i,2)$$

These are called the **2-step probabilities**.
The above formula is exactly the formula for matrix multiplication!

Examples of Markov Chains

Example 2: (a two-state model for weather)

$$\begin{pmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{pmatrix}$$

THEN: 2-step probabilities $\mathrm{p}^2(i,j)$ yield a new 2×2 matrix, and

$$\begin{pmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{pmatrix}^2 = \begin{pmatrix} 0.44 & 0.56 \\ 0.28 & 0.72 \end{pmatrix}$$

i.e. the matrix of 2-step probabilities $\mathrm{p}^2(i,j)$ is simply the **square** of the matrix of transition probabilities $\mathrm{p}(i,j)$.

Examples of Markov Chains

Example 2: (a two-state model for weather)

Question: Out of N days, what proportion of them are rainy?

Observe, the adjoint matrix has an interesting eigenvector

$$\begin{pmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{pmatrix}^t \begin{pmatrix} 1/3 \\ 2/3 \end{pmatrix} = \begin{pmatrix} 1/3 \\ 2/3 \end{pmatrix}$$

The eigenvector corresponds to a probability distribution!.

Later, one of our theorems will guarantee that (with high prob.)

$1/3$ of the N days are rainy

…at least, when N is large.

Transition Probability Matrix

Worth revisiting something we saw earlier: for any i,

$$\sum_j \mathrm{p}(i, j) = 1$$

i.e. the numbers in any row of the transition matrix add to 1.

n-step transition probabilities

If $X_1, X_2, \dots$ denotes a homogeneous Markov chain, then the n-step transition probabilities are defined by

$$\mathrm{p}^n(i, j) = \mathrm{P}(X_n = j \mid X_0 = i)$$

n-step transition probabilities

If $X_1, X_2, \ldots$ denotes a homogeneous Markov chain, then the n-step transition probabilities are defined by

$$\mathrm{p}^n(i,j) = \mathrm{P}(X_n = j \mid X_0 = i)$$

Since the chain is homogeneous, for any k we have

$$\mathrm{P}(X_{k+n} = j \mid X_k = i) = \mathrm{p}^n(i,j)$$

Accordingly, $\mathrm{p}^n(i,j)$ defines a $N \times N$ matrix (where N is the number of states in the system), this is called the n-step transition matrix, which we may denote simply as p^n.

n-step transition probabilities

The Chapman-Kolmogorov equation

Theorem (Chapman-Kolmogorov)

Let $n, m \geq 0$, then, in terms of matrix multiplication, we have

$$\mathrm{p}^{n+m} = \mathrm{p}^n \mathrm{p}^m$$

In particular, the matrix p^n is simply the nth-power of the transition probability matrix.

The Markov Property

Transition Probability

The total probability formula says that

$$\begin{aligned} \mathrm{P}(X_{n+1} = j) &= \sum_k \mathrm{P}(X_{n+1} = j \mid X_n = k)\mathrm{P}(X_n = k) \\ &= \sum_k \mathrm{p}(k, j)\mathrm{P}(X_n = k) \end{aligned}$$

This has a linear algebra interpretation: if we introduce a transition probability matrix with entries $\mathrm{p} = (\mathrm{p}(i, j))_{ij}$, then to go from the distribution of X_n to the distribution of X_{n+1}, all we need to do is multiply the "vector" of probabilities by the **adjoint** of the matrix p.

The Markov Property

Transition Probability

*...all we need to do is multiply the "vector" of probabilities by the **adjoint** of the matrix* p.

This observation leads to what is known as the **Chapman-Kolmogorov** equation, which allows us to compute probabilities of the state several time steps into the future.