Math 456: Mathematical Modeling

Tuesday, February 6th 2018

A few comments from last lecture

1. More on matrix exponential.
2. Gronwall's inequality.
3. Lyapunov functions.

Today:
Discrete time systems and
a rapid course in probability

Tuesday, February 6th 2018

Discrete Time Systems

In preparation for Markov chains and also because they are of interest in their own right we are going to talk about discrete-time dynamical systems.

Here, the continuous time variable t is replaced by a discrete index n. This means that once we are given an **initial state** x_0, we seek to a discrete sequence of states

$$x_1, x_2, \ldots, x_n, x_{n+1}, \ldots$$

representing the evolution of the system. The states could be represented by vectors in some dimension, so

$$x_n \in \mathbb{R}^N$$

Discrete time systems

Instead of a differential equation, we have a recursive relation

$$x_{n+1} = \left(\begin{array}{l} \text{Determined from the previous} \\ \text{or several previous states} \\ \text{via some rule} \end{array} \right)$$

Discrete time systems

Instead of a differential equation, we have a recursive relation

$$x_{n+1} = f_n(x_n, x_{n-1}, x_{n-2}, \ldots, x_{n-k})$$

For some function f, which varies with the time state n we are in, and depending on the current and previous k states of the system.

Discrete time systems

Instead of a differential equation, we have a recursive relation

$$x_{n+1} = f(x_n)$$

A simpler, but big subfamily of problems are those where the f is the same for all n, and it depends only the current state.

Discrete time systems

Instead of a differential equation, we have a recursive relation

$$x_{n+1} = f(x_n)$$

A simpler, but big subfamily of problems are those where the f is the same for all n, and it depends only the current state.

Then, we have to solve

$$\begin{cases} x_{n+1} = f(x_n) \text{ for } n = 0, 1, 2, \ldots, \\ x_0 \text{ is the given initial state.} \end{cases}$$

Remember, the x_n are N-dimensional vectors and $N \geq 1$.

Discrete time systems

Example 1:

Consider the recursive relation (where $N = 1$)

$$x_{n+1} = x_n - 4$$

Can you describe the sequence corresponding to $x_0 = 2$?

Discrete time systems

Example 1:

Consider the recursive relation (where $N = 1$)

$$x_{n+1} = x_n - 4$$

Can you describe the sequence corresponding to $x_0 = 2$?

Answer:

$$x_n = -4n + 2$$

Discrete time systems

Example 2:

Consider now the system (again, $N = 1$)

$$x_{n+1} = 3x_{n-1} + 2$$

Determine x_4 when the system starts at $x_0 = 2$.

Discrete time systems

Example 2:

Consider now the system (again, $N = 1$)

$$x_{n+1} = 3x_{n-1} + 2$$

Determine x_4 when the system starts at $x_0 = 2$.

Answer:

$$\begin{aligned}
x_1 &= 3 \times 2 + 2 = 8 \\
x_2 &= 3 \times 8 + 2 = 26 \\
x_3 &= 78 + 2 = 80 \\
x_4 &= 242
\end{aligned}$$

Discrete time systems

Example 3: (Logistic Growth)

The (discrete) logistic growth system is given by

$$x_{n+1} = \frac{k}{M}\,(M - x_n)\,x_n$$

Here, k and M are positive parameters.

Question: what happens to the system as time goes by, provided x_0 lies between 0 and M?

Discrete time systems

Example 3: (Logistic Growth)

The (discrete) logistic growth system is given by

$$x_{n+1} = \frac{k}{M}\,(M - x_n)\,x_n$$

Here, k and M are positive parameters.

Question: what happens to the system as time goes by, provided x_0 lies between 0 and M?

Answer: x_n increases with n, slowing down as it approaches M, and converging to it in the $n \to \infty$ limit.

Discrete time systems

The previous system has the form

$$x_{n+1} = f(x_n)$$

for a quadratic polynomial f.

The dynamics are fairly simple. This changes, **quite dramatically**, when one goes to two dimensional systems given by quadratic terms.

Discrete time systems

Example 4: (Complex Dynamical Systems and the Julia set)

Think of the two dimensional plane as the complex real numbers, then consider the relation

$$z_{n+1} = z_n^2 + \lambda$$

where $\lambda \in \mathbb{C}$ is some given (complex) parameter.

The **Julia set** (named after Gaston Julia) is the set of initial conditions z_0 such that

$$|z_n| \leq 2 \ \ \forall\, n$$

The Julia set is a region of the plane, let's see how complicated it can get...

Discrete time systems

What you just saw is what may happen when you have dimensions higher than 1 and non-linear terms!!

Let us now consider, instead, higher dimensional systems which are a **linear**. Fix a $N \times N$ matrix A, and let us look at the N dimensional linear system

$$\begin{cases} x_{n+1} = Ax_n \text{ for } n = 0, 1, 2, \ldots \\ x_0 \in \mathbb{R}^N \text{ is given.} \end{cases}$$

Discrete time systems

At the level of the coordinates of x_n, this system looks as follows

$$x_{n+1,1} = \sum_{j=1}^{d} A_{1j} x_{n,j}$$

$$\dots$$

$$x_{n+1,i} = \sum_{j=1}^{d} A_{ij} x_{n,j}$$

$$\dots$$

$$x_{n+1,N} = \sum_{j=1}^{d} A_{Nj} x_{n,j}$$

Discrete time systems

Since all we need to do to advance the system is multiply by A, the n-stage has a relatively simple representation

$$x_n = A^n x_0$$

Then, understanding powers of a matrix is more than enough to understand these systems.

Discrete time systems

Let us illustrate this.

Example 5: (long time decay)
Suppose the matrix A is such that for some $\lambda \in (0,1)$ we have

$$|Ax| \leq \lambda |x| \ \ \forall\, x \in \mathbb{R}^N$$

Then

$$|x_n| \leq \lambda^n |x_0|$$

that is, x_n converges exponentially to 0 as $n \to \infty$.

Next:
A rapid course in probability

A rapid course in probability

We have a set of N states, labeled $\{\omega_1, \ldots, \omega_N\}$
(infinitely many states will be discussed later this semester)

The state of the system is no longer described uniquely, instead the best we can do is determine the probability the system lies in a given state at a given time.

Thus, our state at a given instant is given in terms of a **probability distribution**

$$\mathbb{P}(\text{state is in state } \omega_i) = p_i$$

A rapid course in probability

Example 6:
A six sided die is thrown, each face having equal probability.

$$\mathbb{P}(\{\omega_i\}) = \frac{1}{6} \text{ for } i = 1, \ldots, 6.$$

The probability the number is even,

$$\mathbb{P}(\{\omega_2, \omega_4, \omega_6\}) = \mathbb{P}(\{\omega_2\}) + \mathbb{P}(\{\omega_4\}) + \mathbb{P}(\{\omega_6\}) = \frac{1}{2}$$

A rapid course in probability

Example
A coin has probability $p \in (0,1)$ of heads (represented by 1) and probability $q = 1 - p$ of tails (represented by 0). If one flips this coin 5 times, the probability of the outcome being

$$10001$$

is simply

$$p^2 q^3$$

A rapid course in probability

Example 7:
If I choose one student among two sections, then

$$\mathbb{P}(\{\text{Person ranked SW:TCS above SW:ROT}\}) =$$

A rapid course in probability

So far, we summarize

1. A (finite) probability space is a set of states $\Omega = \{\omega_1, \ldots, \omega_N\}$ combined with a distribution function (or distribution vector), that is, a description of the different probabilities of the states.
2. The probability distribution $\mathbf{p} = (p_1, \ldots, p_N)$ is such that

$$p_i \geq 0 \text{ for } i = 1, \ldots, N; \; p_1 + \ldots + p_N = 1.$$

3. Subsets of Ω are called **events**. The probability of an event A is the sum of the probabilities of the states belonging to that event.

A rapid course in probability

A function on the set of states is known as a **random variable**. Unlike "usual functions", these are denoted with a capital letter. We will mostly look at real random variables

$$X : \{\omega_1, \ldots, \omega_N\} \to \mathbb{R}$$

Random variables give us partial information about our system. They are usually how we experience it.

A rapid course in probability

Expectation and Variance

For a real random variable X, its expectation and variance are the quantities

$$\mathbb{E}[X] := \sum_x x\mathbb{P}[X = x]$$

$$\mathrm{Var}(X) := \mathbb{E}[(X - \mathbb{E}[X])^2]$$

Math 456: Mathematical Modeling

Thursday, February th 2018

Today:
Conditional probability and Gambler's ruin

Thursday, February th 2018

It would make sense to say that an event A is independent from an event B if the **knowledge** that B has occurred does not change the probability that A has occurred as well.

In other words, one **gains no information about** A **from** B.

How does one think about "the probability that A has happened given B did?"

The event "both A and B happen" is simply $A \cap B$.

Accordingly, the probability of A and B both happening is

$$\mathbb{P}(A \cap B)$$

Since now the event B happened, we need to reweight our probability so that B itself is assign probability 1.

Therefore, it makes sense to think of the quantity

$$\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

as the probability of A **given** B.

A different name for this quanity is **the conditional probability of A with respect to B**, denoted by

$$\mathbb{P}(A \mid B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

If one reverses the roles of A and B, one obtains

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$$

Which leads to the interesting relation

$$\mathbb{P}(B)\mathbb{P}(A \mid B) = \mathbb{P}(A)\mathbb{P}(B \mid A)$$

This useful, if rather elementary observation, is better known as

"Bayes Theorem"

and it is stated in the following form:

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A)\mathbb{P}(B \mid A)}{\mathbb{P}(B)}$$

Going back to our earlier comment

*It would make sense to say that an event A is independent from an event B if the **knowledge** that B has occurred does not change the probability that A has occurred as well.*

we see that we should say that A **is independent** of B when

$$\mathbb{P}(A \mid B) = \mathbb{P}(A)$$

Bayes Theorem, or even just the very definition $\mathbb{P}(A \mid B)$ makes it clear that A being independent of B is the same as B being independent of A.

Independence is then a symmetric relationship between events, and it can be rewritten as

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

Divide and conquer:

Given disjoint events $A_1, A_2, \ldots, A_N$ and another set E, we can compute the probability of E if we know the probability of E given any one of the events A_k, and their probbabilities

Divide and conquer:

Given disjoint events $A_1, A_2, \ldots, A_N$ and another set E, we can compute the probability of E if we know the probability of E given any one of the events A_k, and their probbabilities

$$\mathbb{P}(E) = \mathbb{P}(E \mid A_1)\mathbb{P}(A_1) + \ldots + \mathbb{P}(E \mid A_N)\mathbb{P}(A_N)$$

This is often called the **total probability formula**.

Example 1:
Let us throw a symmetric, 6-sided die twice, and let us denote by X_1 and X_2 and the numbers from these two throws.

The outcome from these two throws are independent from one another.

Question: What is the probability that $X_1 + X_2 = 3$?

Divide and conquer (II):

A different approach involves sequence of events that *become more and more concrete*. If you have a set of events $A_1, \ldots, A_k$, and they are nested, that is if

$$A_k \subset A_{k-1} \subset A_{k-2} \subset \ldots \subset A_1$$

Then, one can compute the probability of A_k if you know the probability of A_1, as well as the succesive conditional probabilities

$$\mathbb{P}(A_k) = \mathbb{P}(A_1)\mathbb{P}(A_2 \mid A_1)\mathbb{P}(A_3 \mid A_2) \ldots \mathbb{P}(A_k \mid A_{k-1})$$

A rapid course in probability

Let us revisit an example from last class...

Example 2:
A coin has probability $p \in (0, 1)$ of heads (represented by 1) and probability $q = 1 - p$ of tails (represented by 0). If one flips this coin 2 times, the probability of the outcome being

$$10$$

is simply

$$pq$$

A rapid course in probability

Let us revisit an example from last class...

Example 2:
Why is this so?

$$10$$

Well, this is the probability of two events happening, and we are modeling these two events as having nothing to do with one another...

A rapid course in probability

Let us revisit an example from last class...

Example 2:
Therefore

$$\begin{aligned}&\mathbb{P}(\text{second flip is tails} \mid \text{first flip is heads})\\&= \mathbb{P}(\text{scond flip is tails}) = q\end{aligned}$$

Then,

$$\mathbb{P}(\text{first flip is heads, second is tails }) = pq$$

A rapid course in probability

Let us revisit an example from last class...

Example 2:
In particular, the probability of 5 flips resulting in the sequence

$$10001$$

is nothing but

$$p^2 q^3$$

A rapid course in probability

What the previous examples show is this:

Conditional probabilities give us a way to write down, that is, **to propose**, more complicated probability distributions from the probabilities of simpler events.

Example 3:

You are playing a simple 2-D video game, and the the enemy "AI" works as follows: the code chooses among any of the neighboring sites, at random, and moves there, all neighbors being equally likely.

What is the probability that if at time $n = 0$ the enemy starts as indicated in the sketch, that at $n = 3$ it lies at the end point?

A rapid course in probability

Example 4:
You currently have 100 dollars, you and a friend are taking bets on the world series, for every game the Yankees win you pay your friend 10 dollars, for every game the Diamondbacks win your friend pays you 10 dollars.

Under the hypothesis that the probability that the Yankees beat the Diamondbacks 6 times out of 10, what is the probability you have exactly 120 dollars after 3 games?

A rapid course in probability

Independence of random variables

One can also talk about independence for random variables.

Two random variables Y_1, Y_2 are said to be **independent** if

$$\mathbb{P}(Y_1 = a, Y_2 = b) = \mathbb{P}(Y_1 = a)P(Y_2 = b)$$

for all values a and b.

... if there are Y_1, Y_2, Y_3, **independence** means that

$$\mathbb{P}(Y_1 = a, Y_2 = b, Y_3 = c) = \mathbb{P}(Y_1 = a)P(Y_2 = b)P(Y_3 = c)$$

and so on for larger, finite number of variables.

A rapid course in probability

Independence of random variables

... if there are Y_1, Y_2, Y_3, **independence** means that

$$\mathbb{P}(Y_1 = a, Y_2 = b, Y_3 = c) = \mathbb{P}(Y_1 = a)P(Y_2 = b)P(Y_3 = c)$$

and so on for larger, finite number of variables.

If one is given an infinite sequence $Y_1, Y_2, \ldots, Y_n, \ldots$ then they are independent if every finite subset of them is independent.

Durrett's textbook:
From this point on and for the next few two or three classes, we will be covering material in Sections 1.1 and 1.2. from the book.

Gambler's Ruin

Sum of independent Bernoulli Random Variables

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

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

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

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

Let's compute this!

Gambler's Ruin

Sum of independent Bernoulli Random Variables

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

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

Gambler's Ruin

Sum of independent Bernoulli Random Variables

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

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

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

Gambler's Ruin

Sum of independent Bernoulli Random Variables

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

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

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

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

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

Gambler's Ruin

Sum of independent Bernoulli Random Variables

In conclusion,

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

Moreover,

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

Gambler's Ruin

Sum of independent Bernoulli Random Variables

In conclusion,

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

Moreover,

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

THUS: **Knowledge** of X_n **simplifies** the distribution of X_{n+1}.