Winter 2020 Math 126
Topics in Applied Mathematics
Data-driven Uncertainty Quantification

Yoonsang Lee (yoonsang.lee@dartmouth.edu)

Lecture 1: Introduction

'19 Winter M126

- Class: MWF 12:50-1:55 pm @ 200 Kemeny
- X-hour: Tuesday 1:20-2:10 pm
- Office hours: WF 2-3 pm or by appointment
- Instructor: Yoonsang Lee (yoonsang.lee@dartmouth.edu)
- Office: 206 Kemeny
- Grading: homework (40%), midterm (20%), final exam (40%)

- Homework will include theory and computer simulation problems
- More details including lecture notes are available at http://math.dartmouth.edu/∼m126w20

'19 Winter M126

- Class: MWF 12:50-1:55 pm @ 200 Kemeny
- X-hour: Tuesday 1:20-2:10 pm
- Office hours: WF 2-3 pm or by appointment
- Instructor: Yoonsang Lee (yoonsang.lee@dartmouth.edu)
- Office: 206 Kemeny
- Grading: homework (40%), midterm (20%), final exam (40%) **if** there are more than 5 first-year graduate or undergraduate students enrolled in the course.
- Homework will include theory and computer simulation problems
- More details including lecture notes are available at http://math.dartmouth.edu/∼m126w20

Uncertainty Quantification (UQ)

- Deterministic models of complex physical processes incorporate some element of **uncertainty** to account for **lack of knowledge** about important physical parameters, random variability, or ignorance about what the form of a 'correct' model would be..
- Goal of UQ: to provide accurate predictions about systems' uncertainty behavior.
- Roughly speaking, UQ is the interplay of probability theory and statistical practice with 'the real world' applications'
- This course will cover more than probability/statistics.

An example of uncertainty in a deterministic system

- Let x_n be the value of a physical system at the n-th time.
- We want to know the future value of x_n, say x_{n+1}.
- Using fancy and esoteric physics/mathematics techniques, we found that the future value is given by

$$x_{n+1} = \text{fractional part of } 10x_n$$

Q : How many times can we repeat for accurate prediction?

An example of uncertainty in a deterministic system

Let's assume that the true initial value is

$$x_0 = 0.123456789123456789....$$

Due to uncertainty in getting the correct initial value, let us assume that we start with an initial value $\tilde{x}_0 = 0.123$.

An example of uncertainty in a deterministic system

Let's assume that the true initial value is

$$x_0 = 0.123456789123456789....$$

Due to uncertainty in getting the correct initial value, let us assume that we start with an initial value $\tilde{x}_0 = 0.123$.

- $x_{n+1} =$ fractional part of $10x_n$
  $x_1 = 0.234567891...$, $x_2 = 0.345678912...$, $x_3 = 0.456789123...$
  $\tilde{x}_1 = 0.23$, $\tilde{x}_2 = 0.30$, $\tilde{x}_3 = 0.00$

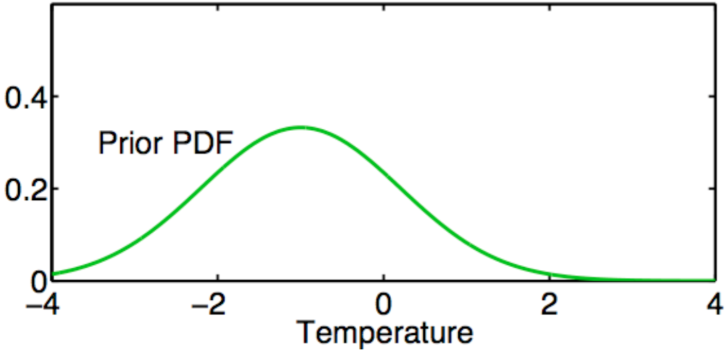
No prediction!

Data-driven Uncertainty Quantification

- Data provides information of the uncertain information.
- Goal of data-driven UQ: to improve predictions using data.
- The information included in data is typically incomplete and noisy.

An example of improving prediction using data

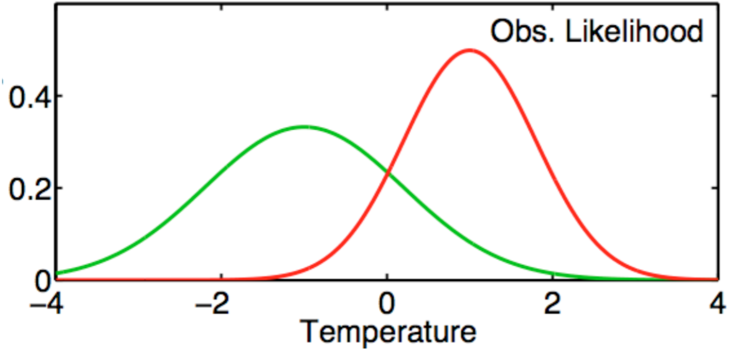
Prediction of temperature



PDF: probability density function

An example of improving prediction using data

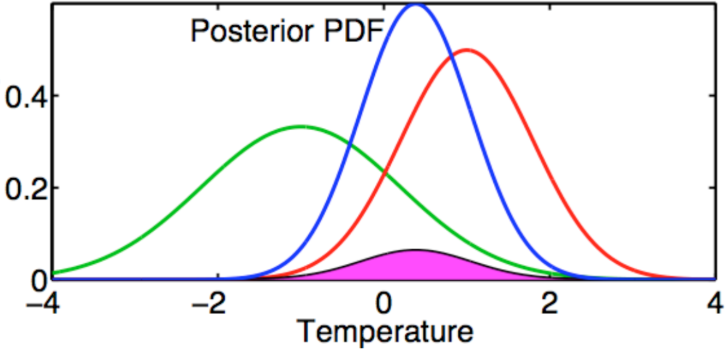
Prediction of temperature



Observation data is noisy; the variance of the red PDF is the observation error variance.

An example of improving prediction using data

Prediction of temperature



The posterior PDF is obtained by the **Bayes' formula**

$$p(t|v) \approx p(t)p(v|t), \qquad t : \text{temperature}, v : \text{observation}$$

Is it raining?

(Facebook data scientist interview question)

You're about to get on a plane to Seattle. You want to know if you should bring an umbrella. You call 3 random friends of yours who live there and ask each independently if it's raining. Each of your friends has a 2/3 chance of telling you the truth and a 1/3 chance of messing with you by lying. All 3 friends tell you that "Yes" it is raining. What is the probability that it's actually raining in Seattle?

Is it raining?

(Facebook data scientist interview question)

You're about to get on a plane to Seattle. You want to know if you should bring an umbrella. You call 3 random friends of yours who live there and ask each independently if it's raining. Each of your friends has a 2/3 chance of telling you the truth and a 1/3 chance of messing with you by lying. All 3 friends tell you that "Yes" it is raining. What is the probability that it's actually raining in Seattle?

Hint if you have more responses, are you more confident?

Is it raining?

(Facebook data scientist interview question)

You're about to get on a plane to Seattle. You want to know if you should bring an umbrella. You call 3 random friends of yours who live there and ask each independently if it's raining. Each of your friends has a 2/3 chance of telling you the truth and a 1/3 chance of messing with you by lying. All 3 friends tell you that "Yes" it is raining. What is the probability that it's actually raining in Seattle?

Hint if you have more responses, are you more confident?

Answer

$$\begin{aligned} P(rain|y,y,y) =& \frac{P(y,y,y|rain)P(rain)}{P(y,y,y)} \quad \text{using Bayes' theorem} \\ =& \frac{P(y,y,y|rain)P(rain)}{P(y,y,y|rain)P(rain)+P(y,y,y|no\_rain)P(no\_rain)} \\ =& \frac{(2/3)^3 P(rain)}{(2/3)^3 P(rain)+(1/3)^3(1-P(rain))} \end{aligned} \tag{1}$$

Big picture of the course

Please keep the following questions in mind during the course

- How to (effectively) represent uncertainty?
- How to propagate uncertainty in time (for future predictions)?
- How to incorporate data?
- How to extract useful information?

Big picture of the course

Please keep the following questions in mind during the course

- How to (effectively) represent uncertainty?
  What is the best strategy to represent a non-Gaussian distribution? What about high-dimensional spaces?
- How to propagate uncertainty in time (for future predictions)?
  Given the current PDF, how can we obtain the future time PDF? What about high-dimensional spaces?
- How to incorporate data?
  What if we have only a small number of data? What if observation is not directly related to the uncertain variable?
- How to extract useful information?

Specific goals of the course

- This course will focus on intuition behind UQ methods rather than technical details.
- Although I do not plan to teach advanced mathematics, I will motivate you why we need them for applied and computational mathematics.
- You will be familiar with/able to run computational UQ methods for data-driven sciences/engineering.

Plan (subject to changes)

Week 1: Introduction

- Day 1 - Intro and overview
- Day 2 - Review of probability
- Day 3 - Information theory

Week 2: Statistical Inference

- Day 1 - Parametric inference
- Day 2 - Bayesian inference
- Day 3 - Nonparametric estimation

Week 3: Random Sampling

- Day 1 - Monte Carlo
- Day 2 - Markov chain Monte Carlo
- Day 3 - Importance sampling

Plan (subject to changes)

Week 4: Polynomial Chaos

- Day 1 - Hilbert Space
- Day 2 - Polynomial Chaos
- Day 3 - Stochastic Galerkin

Week 5: Special Tools

- Day 1 - Optimization (variational techniques)
- Day 2 - Smoothing using orthogonal functions
- Day 3 - Midterm

Week 6: Random Systems

- Day 1 - Chaotic Systems
- Day 2 - Stochastic differential equations
- Day 3 - Multiscale and long-time behavior

Plan (subject to changes)

Week 7: Data Assimilation

- Day 1 - Kalman Filter
- Day 2 - 3D-Var
- Day 3 - Ensemble Kalman Filter

Week 8: Advanced Data Assimilation

- Day 1 - Ensemble square root filter
- Day 2 - Particle filter
- Day 3 - Localization and inflation

Week 9: Challenges of High-dimensional Spaces

- Day 1 - Sampling in high-dimensional spaces
- Day 2 - Data assimilation in high-dimensional spaces
- Day 3 - Review

Sample homework/test problems

- Do online search for 'Kalman Filter' and make a one-sentence summary.
- For a variable u and its observation v, is the conditional probability $p(u|v)$ larger than the prior $p(u)$? In other words, does the observation v decrease the uncertainty of u?
- Modify the given data assimilation code to incorporate the deterministic sample-transformation method you learned.

Special minor topic I: Why log? - entropy

Assume that we have a system of N particles with volume V and energy E. Let Ω represent all possible configurations of the system. From the elementary Physics, we all know that the entropy $\mathcal{S}(\Omega)$ is defined by

$$\mathcal{S} = -k \ln \Omega,$$

where k is the Boltzmann constant.

Special minor topic I: Why log? - entropy

We want to define entropy $\mathcal{S}(\Omega)$ to represent uncertainty of the system. Entropy must satisfy

- From two systems of Ω_1 and Ω_2, the total entropy of the combined system, S_{12} (not necessarily in contact), is given by

$$\mathcal{S}_{12} = \mathcal{S}_1 + \mathcal{S}_2$$

or

$$\mathcal{S}(\Omega_{12}) = \mathcal{S}(\Omega_1) + \mathcal{S}(\Omega_2)$$

The relation between Ω_{12}, Ω_1 and Ω_2

$$\Omega_{12} = \Omega_1 \times \Omega_2$$

Special minor topic I: Why log? - entropy

What is the general form of functions satisfying

$$\mathcal{S}(\Omega_1\Omega_2) = \mathcal{S}(\Omega_1) + \mathcal{S}(\Omega_2)$$

Special minor topic I: Why log? - entropy

What is the general form of functions satisfying

$$\mathcal{S}(\Omega_1 \Omega_2) = \mathcal{S}(\Omega_1) + \mathcal{S}(\Omega_2)$$

Answer: $\mathcal{S}(\Omega) = k \ln \Omega \quad$ for a constant k

Special minor topic I: Why log? - entropy

Another interpretation of entropy

The probability of the system being in one of the all possible states (Ω)

$$p = \frac{1}{\Omega}$$

Special minor topic I: Why log? - entropy

Another interpretation of entropy

The probability of the system being in one of the all possible states (Ω)

$$p = \frac{1}{\Omega}$$

$$\mathcal{S} = \ln \Omega = -\ln p$$

Special minor topic I: Why log? - entropy

Another interpretation of entropy
The probability of the system being in one of the all possible states (Ω)

$$p = \frac{1}{\Omega}$$

$$\mathcal{S} = \ln \Omega = -\ln p$$

An example in communication We want to send a number from 0 to 1023. The probability of receiving, for example, 263, is $\frac{1}{1024}$. How much information we need to identify the number? That is, what is the uncertainty of this communication?

Special minor topic I: Why log? - entropy

Another interpretation of entropy
The probability of the system being in one of the all possible states (Ω)

$$p = \frac{1}{\Omega}$$

$$\mathcal{S} = \ln \Omega = -\ln p$$

An example in communication We want to send a number from 0 to 1023. The probability of receiving, for example, 263, is $\frac{1}{1024}$. How much information we need to identify the number? That is, what is the uncertainty of this communication?

$$-\log_2 \frac{1}{1024} = 10\text{bits}$$

Special minor topic I: Why log? - entropy

Assume that there are only M states with corresponding probabilities $p_1, p_2, ..., p_M$. The entropy of each state is

$$\mathcal{S}^m = -\ln p_m$$

What is the average entropy?

Special minor topic I: Why log? - entropy

Assume that there are only M states with corresponding probabilities $p_1, p_2, ..., p_M$. The entropy of each state is

$$\mathcal{S}^m = -\ln p_m$$

What is the average entropy?

$$\boxed{H(\{p_m\}) = \sum_{m=1}^{M} p_m \mathcal{S}^m = \sum -p_m \ln p_m}$$

Question 1 We have a system that can take M different states. What is the probability maximizing the entropy?

Question 1 We have a system that can take M different states. What is the probability maximizing the entropy?

Answer Find the max of $H(p)$? No. We need to consider the constraint

$$\sum_m p(x_m) = \sum_m p_m = 1$$

Question 1 We have a system that can take M different states. What is the probability maximizing the entropy?

Answer Find the max of $H(p)$? No. We need to consider the constraint

$$\sum_m p(x_m) = \sum_m p_m = 1$$

Find the max of
$\tilde{H}(p) := H(p) + \lambda(\sum p_m - 1) = -\sum p_m \ln p_m + \lambda(\sum p_m - 1)$

$$\partial \tilde{H}/\partial p_m = -\ln p_m - 1 + \lambda = 0$$

$$p_m \approx \exp(-1+\lambda) \quad \text{constant}$$

Question 2 For the same system, we know from an experiment that the total energy is constant, $\overline{E} = \sum E_m p_m$. What is the probability maximizing the entropy?

Question 2 For the same system, we know from an experiment that the total energy is constant, $\overline{E} = \sum E_m p_m$. What is the probability maximizing the entropy?

Answer

$$\tilde{H}(p) = -\sum p_m \ln p_m + \lambda_1(\sum p_m - 1) + \lambda_2(\sum E_m p_m - \overline{E})$$

Question 2 For the same system, we know from an experiment that the total energy is constant, $\overline{E} = \sum E_m p_m$. What is the probability maximizing the entropy?

Answer

$$\tilde{H}(p) = -\sum p_m \ln p_m + \lambda_1\left(\sum p_m - 1\right) + \lambda_2\left(\sum E_m p_m - \overline{E}\right)$$

$$\partial \tilde{H}/\partial p_m = -\ln p_m - 1 + \lambda_1 + \lambda_2 E_m = 0$$

$$p_m \approx \exp(\lambda_2 E_m)$$

Question 2 For the same system, we know from an experiment that the total energy is constant, $\overline{E} = \sum E_m p_m$. What is the probability maximizing the entropy?

Answer

$$\tilde{H}(p) = -\sum p_m \ln p_m + \lambda_1(\sum p_m - 1) + \lambda_2(\sum E_m p_m - \overline{E})$$

$$\partial \tilde{H}/\partial p_m = -\ln p_m - 1 + \lambda_1 + \lambda_2 E_m = 0$$

$$p_m \approx \exp(\lambda_2 E_m)$$

The **Boltzmann** distribution for physicists (or the **Gibbs** distribution for mathematicians)!

Question 2 For the same system, we know from an experiment that the total energy is constant, $\overline{E} = \sum E_m p_m$. What is the probability maximizing the entropy?

Answer

$$\tilde{H}(p) = -\sum p_m \ln p_m + \lambda_1(\sum p_m - 1) + \lambda_2(\sum E_m p_m - \overline{E})$$

$$\partial \tilde{H} / \partial p_m = -\ln p_m - 1 + \lambda_1 + \lambda_2 E_m = 0$$

$$p_m \approx \exp(\lambda_2 E_m)$$

The **Boltzmann** distribution for physicists (or the **Gibbs** distribution for mathematicians)!

Homework Search for the Boltzmann (or Gibbs) distribution and make a one-sentence summary.

Special minor topic II: functional derivative

Find the derivative of $\int_0^1 (\frac{df}{dx})^2 dx$ with respect to f

Special minor topic II: functional derivative

Find the derivative of $\int_0^1 (\frac{df}{dx})^2 dx$ with respect to f

$$\int_0^1 2\frac{df}{dx}dx = 2f(1) - 2f(0)$$

Special minor topic II: functional derivative

Find the derivative of $\int_0^1 (\frac{df}{dx})^2 dx$ with respect to f

$$\int_0^1 2\frac{df}{dx} dx = 2f(1) - 2f(0)$$

Incorrect!

$$\begin{aligned} &\int_0^1 \left((\tfrac{df}{dx} + \tfrac{d\delta}{dx})^2 - (\tfrac{df}{dx})^2 \right) dx \\ = &\int_0^1 \left((\tfrac{df}{dx})^2 + 2\tfrac{df}{dx}\tfrac{d\delta}{dx} + (\tfrac{d\delta}{dx})^2 \right) dx \\ = &\int_0^1 2\tfrac{df}{dx}\tfrac{d\delta}{dx} dx \\ = &-\int_0^1 2\tfrac{d^2f}{dx^2}\delta dx \end{aligned}$$

δ is arbitrary; thus if we are looking for an equilibrium point (gradient is zero), we must have

$$-2\frac{d^2f}{dx^2} = 0$$

Special minor topic II: functional derivative

Schrodinger equation

One-dimensional Schrodinger equation with a potential V:

$$-\psi_{xx} + V\psi = 0$$

ψ : wave function, $|\psi|^2$: probability distribution function

Find ψ minimizing $\int \frac{1}{2}|\psi_x|^2 + V(x)|\psi|^2 dx$

Special minor topic II: functional derivative

Schrodinger equation

One-dimensional Schrodinger equation with a potential V:

$$-\psi_{xx} + V\psi = 0$$

ψ : wave function, $|\psi|^2$: probability distribution function

Find ψ minimizing $\int \frac{1}{2}|\psi_x|^2 + V(x)|\psi|^2 dx$

$$\Rightarrow -\psi_{xx} + V\psi = 0$$

Special minor topic II: functional derivative

Schrodinger equation

One-dimensional Schrodinger equation with a potential V:

$$-\psi_{xx} + V\psi = 0$$

ψ : wave function, $|\psi|^2$: probability distribution function

Find ψ minimizing $\int \frac{1}{2}|\psi_x|^2 + V(x)|\psi|^2 dx$

$$\Rightarrow -\psi_{xx} + V\psi = 0$$

Do you know other examples (equations), which contain second derivatives?

Now it is time to use the special topics I and II

Question: What is the probability distribution function p maximizing the entropy with a fixed mean $0 = \overline{x} = \int xp(x)dx$ and variance $\sigma^2 = \int (x - \overline{x})^2 p(x)dx$?

Now it is time to use the special topics I and II

Question: What is the probability distribution function p maximizing the entropy with a fixed mean $0 = \overline{x} = \int xp(x)dx$ and variance $\sigma^2 = \int (x - \overline{x})^2 p(x)dx$?

$$\begin{aligned}\tilde{H} &= -\int p \ln p dx + \lambda_1(\int p dx - 1) + \lambda_2 \int xp dx \\ &\quad + \lambda_3(\int x^2 p dx - \sigma^2) \end{aligned} \tag{2}$$

$$\partial \tilde{H}/\partial p = -\int (\ln p + 1)dx + \lambda_1 \int 1 dx + \lambda_2 \int x dx + \lambda_3 \int x^2 dx$$

Now it is time to use the special topics I and II

Question: What is the probability distribution function p maximizing the entropy with a fixed mean $0 = \overline{x} = \int xp(x)dx$ and variance $\sigma^2 = \int (x - \overline{x})^2 p(x)dx$?

$$\begin{aligned} \tilde{H} &= -\int p \ln p dx + \lambda_1 (\int p dx - 1) + \lambda_2 \int x p dx \\ &\quad + \lambda_3 (\int x^2 p dx - \sigma^2) \end{aligned} \tag{2}$$

$$\partial \tilde{H}/\partial p = -\int (\ln p + 1) dx + \lambda_1 \int 1 dx + \lambda_2 \int x dx + \lambda_3 \int x^2 dx$$

$p(x) \approx \exp(x^2/2\sigma^2)$ **Gaussian** distribution!!!

Programming

- Computational methods are essential for UQ as analytic tools are limited. Thus this course will focus on implementation and validation of UQ methods for applications. You are welcome to use any programming langue of your preference including MATLAB.
- If you are interested in experiences other than the school setting (for example, internship at industrial companies or national labs), it is highly recommended to use Python/C++ for your programming.

Programming

- Computational methods are essential for UQ as analytic tools are limited. Thus this course will focus on implementation and validation of UQ methods for applications. You are welcome to use any programming langue of your preference including MATLAB.
- If you are interested in experiences other than the school setting (for example, internship at industrial companies or national labs), it is highly recommended to use Python/C++ for your programming.

Homework Open an account on Discovery, the Dartmouth HPC cluster, and learn how to run MATLAB/python codes.