Winter 2021 Math 106
Topics in Applied Mathematics
Data-driven Uncertainty Quantification

Yoonsang Lee (yoonsang.lee@dartmouth.edu)

Lecture 2: Review of Probability

Review of Lecture 1

- Course Webpage: `http://math.dartmouth.edu/~m106w21`
- Bayes' theorem

$$p(u|v) \approx p(u)p(v|u)$$

- (Average) entropy

$$H(\{p\}) = -\sum_{m}^{M} p_m \ln p_m$$

The maximum entropy distribution (or equilibrium distribution) of fixed mean and variance is given by the Gaussian distribution.

Probability

Probability plays an important role in UQ. We will review some basic facts of probability in this lecture.

2.1 Probability space $(\Omega, \mathcal{B}, \mu)$

The triple $(\Omega, \mathcal{B}, \mu)$ is called a *probability space* where

Def. A sample space Ω is the space of all possible outcomes.

Def. $\mathcal{B}$ is a σ-algebra if it satisfies the following properties

1. $\emptyset \in \mathcal{B}$ and $\Omega \in \mathcal{B}$
2. If $B \in \mathcal{B}$, then its complement $B^c = \Omega \backslash B \in \mathcal{B}$.
3. For $\{A_i, i \in \mathbb{N}\}$, then $\bigcup_i A_i \in \mathcal{B}$

Def. A probability measure $\mu(A)$ for $A \in \mathcal{B}$ is a function $\mu : \mathcal{B} \to \mathbb{R}$ such that

1. $\mu(\Omega) = 1$
2. $0 \leq \mu \leq 1$.
3. If $\{A_1, A_2, ..., A_n, ...\}$ is a finite or countable collection of events such that $A_i \in \mathcal{B}$ and $A_i \cap A_j = \emptyset$ for $i \neq j$, $\mu(\bigcup_i^\infty A_i) = \sum_i^\infty \mu(A_i)$

2.1 Probability space $(\Omega, \mathcal{B}, \mu)$

Def. An element ω of Ω is an outcome.
Def. An element element of $\mathcal{B}$ is called an event.
Def. A random variable $X : \Omega \to \mathbb{R}$ is a $\mathcal{B}$-measurable function defined on Ω, where $\mathcal{B}$-measurable means that the subset of elements $\omega \in \Omega$ for which $X(\omega) \leq x$ is an element of $\mathcal{B}$ for every $x \in \mathbb{R}$.
Def. Given a probability measure $\mu(A)$, the probability *distribution* function of a random variable X, P_XA, is defined by

$$P_X(x) = \mu(X \leq x)$$

Def. If P_X is differentiable, its derivative, $p(x) = P'_X(x)$ is called the probability *density* of X.

2.1.1 Examples of probability densities

- Bernoulli density. Let X represent a binary coin flip with $\mu(X = 1) = p$ and $\mu(X = 0) = 1 - p$ for some $p \in [0, 1]$. The probability density is

$$p(x) = p^x(1 - p)^{1-x} \text{ for } x \in \{0, 1\}.$$

- Binomial density. Flip the above coin n times and let X be the number of heads. Assume that the tosses are independent. For $x = 1, 2, ..., n$,

$$p(x) = \binom{n}{x} p^x(1 - p)^{n-x}$$

2.1.1 Examples of probability densities

- Bernoulli density. Let X represent a binary coin flip with $\mu(X = 1) = p$ and $\mu(X = 0) = 1 - p$ for some $p \in [0, 1]$. The probability density is

$$p(x) = p^x(1 - p)^{1-x} \text{ for } x \in \{0, 1\}.$$

- Binomial density. Flip the above coin n times and let X be the number of heads. Assume that the tosses are independent. For $x = 1, 2, ..., n$,

$$p(x) = \binom{n}{x} p^x(1 - p)^{n-x}$$

Exercise What is the sample space of the Bernoulli distribution? What is is corresponding probability measure?

2.1.1 Examples of probability densities

- Gaussian (or normal) density with mean m and variance σ^2, $N(m, \sigma^2)$

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

,

- Uniform density on the interval (a, b)

$$p(x) = \begin{cases} \frac{1}{b-a}, & x \in (a, b), \\ 0, & x \notin (a, b) \end{cases}$$

- Cauchy density

$$p(x) = \frac{1}{\pi(1 + x^2)}$$

2.1.2 Transformations of random variables

Let X and Y be two random variables and r is a relation between them, that is, $Y = r(X)$. If $p(x)$ is the density of X, what is the density of Y, say $f(y)$ in terms of p and y?

2.1.2 Transformations of random variables

Let X and Y be two random variables and r is a relation between them, that is, $Y = r(X)$. If $p(x)$ is the density of X, what is the density of Y, say $f(y)$ in terms of p and y?

Answer

When r is monotone and differentiable,

$$p(x)dx = p(r^{-1}(y))|\frac{dr^{-1}}{dy}|dy$$

$$\text{Thus, } f(y) = p(r^{-1}(y)|\frac{dr^{-1}}{dy}|$$

2.1.2 Transformations of random variables

Let X and Y be two random variables and r is a relation between them, that is, $Y = r(X)$. If $p(x)$ is the density of X, what is the density of Y, say $f(y)$ in terms of p and y?

Example

Let $p(x) = e^{-x}$ for $x > 0$ and $Y = r(x) = \log X$. From the change of variables,

$$f(y) = p(e^y)\frac{de^y}{dy} = e^{-e^y}e^y$$

2.1.2 Transformations of random variables

Let X and Y be two random variables and r is a relation between them, that is, $Y = r(X)$. If $p(x)$ is the density of X, what is the density of Y, say $f(y)$ in terms of p and y?

Answer

In general case, use the following steps

1. For each y, find the set $A_y = \{x | r(x) \leq y\}$.
2. $P_Y(y) = \mu(Y \leq y) = \mu(r(X) \leq y) = \mu(\{x | r(x) \leq y\}) = \int_{A_y} p(x)dx$
3. $f(y) = P'_y(y)$.

2.1.2 Transformations of random variables

Let X and Y be two random variables and r is a relation between them, that is, $Y = r(X)$. If $p(x)$ is the density of X, what is the density of Y, say $f(y)$ in terms of p and y?

Example

Let $p(x) = e^{-x}$ for $x > 0$ and $Y = r(x) = \log X$. Then, $P_X(x) = \int_0^x p(t)dt = 1 - e^{-x}$ and $A_y = \{x|x \leq e^y\}$.

$$P_Y(y) = \mu(Y \leq y) = \mu(\log X \leq y) = \mu(X \leq e^y) = P_X(e^y) = 1 - e^{-e^y}.$$

Therefore, $f(y) = e^y e^{-e^y}$.

2.1.2 Transformations of random variables

Exercise X is uniform on $[0, 2\pi]$. Find the density of $Y = \sin X$.

Exercise Let X_1 and X_2 are two independent uniform distributions on $(0, 1)$.

1. Find the density of $Y_1 = X_1 + X_2$.
2. Find the density of $Y_2 = X_1 - X_2$.
3. Find the density of $Y_3 = X_1/X_2$.
4. Find the density of $Y_4 = \max(X_1, X_2)$.

2.2 Expected Values and Moments

Def. Let $(\Omega, \mathcal{B}, \mu)$ be a probability space and X a random variable. Then the expected value (or mean) of the random variable X is defined as the integral of X over Ω with respect to the measure μ

$$E[X] = \int_{\Omega} X(\omega) d\mu = \int xp(x)dx.$$

Def. The variance $Var(X)$ of the random variable X is

$$Var(X) = E[(X - E[X])^2] = \int (x - E[X])^2 p(x) dx$$

and the standard deviation of X is

$$\sigma(X) = \sqrt{Var(X)}.$$

2.2 Expected Values and Moments

Def. The m-th moment of a random variable X is defined by

$$E[X^m] = \int_{-\infty}^{\infty} x^m p(x) dx$$

2.2 Expected Values and Moments

Def. The m-th moment of a random variable X is defined by

$$E[X^m] = \int_{-\infty}^{\infty} x^m p(x) dx$$

Thm. If the m-th moment exists and $j < m$ then the j-th moment exists.

Proof.

$$\begin{aligned} E[X^j] = \int_{-\infty}^{\infty} x^j p(x) dx &= \int_{|x|\leq 1} x^j p(x) dx + \int_{|x|>1} x^j p(x) dx \\ &\leq \int_{|x|\leq 1} + \int_{|x|>1} x^j p(x) dx \\ &\leq 1 + E[X^m] < \infty. \end{aligned}$$

2.2 Expected Values and Moments

Exercise

- Find the mean and variance of a Gaussian random variable X with a density $p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-(x-m)^2/2\sigma^2)$.
- Find the mean of the Cauchy distribution $p(x) = \frac{1}{\pi(1+x^2)}$.
- Find the mean and variance of the Binomial distribution $b(x; n, p) = \binom{n}{x} p^x (1-p)^{n-x}$.

2.2 Expected Values and Moments

Exercise Let X be a random variable such that $E[|X|^m] \leq AC^m$ for some positive constancts A and C, and all intergers $m \geq 0$. Show that $\mu(|X| > C) = 0$.

2.3 Joint Probability and Independence

Def. Two events A and B, $A, B \in \mathcal{B}$, are *independent* if $\mu(A \cap B) = \mu(A)\mu(B)$.

Def. Two random variables X and Y are *independent* if the events $\{X \leq x\}$ and $\{Y \leq y\}$ are independent for all x and y.

Def. The joint distribution of two random variables X and Y is defined by

$$P_{XY}(x, y) = \mu(X \leq x, Y \leq y)$$

Def. If the second mixed derivative $\partial^2 P_{XY}(x, y)/\partial x \partial y$ exists, it is called the joint probability density

$$P_{XY}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} p(s, t) dt ds$$

2.3 Joint Probability and Independence

Def. The covariance of two random variables X and Y is

$$Cov(X, Y) = E[(X - E[X])(Y - E[Y])].$$

Def. Correlation between X and Y is defined as

$$Cor(X, Y) = \frac{Cov(X, Y)}{\sigma(X)\sigma(Y)}$$

Def. Two random variables X and Y are uncorrelated if $Cor(X, Y) = 0$.

Note. X and Y are independent $\Rightarrow$ X and Y are uncorrelated. The opposite direction does not hold.

Def. The marginal densities of X and Y are

$$p(x) = \int p(x, y)dy, \quad p(y) = \int p(x, y)dx$$

2.3 Joint Probability and Independence

Exercise (programming) Generate a sample of two random variables X and Y where X and Y are normal with a correlation ρ.

2.4 Conditional Probability and Conditional Expectation

Def. The probability of an event B given an event A is defined by

$$\mu(A|B) = \frac{\mu(A \cap B)}{\mu(B)}.$$

Def. If two random variables X and Y have densities p_X and p_Y respectively, the conditional probability density of X given Y is defined by

$$p_{Y|X}(y|x) = \frac{p_{X,Y}(x,y)}{p_X(x)}$$

Def. The conditional expectation of X given Y is defined by

$$E[Y|X] = \int y p_{Y|X}(y|x) dx$$

2.4 Conditional Probability and Conditional Expectation

Exercise Let X and Y be two random variables with $E[Y] = m$ and $E[Y^2] < \infty$.

1. Show that the constant c that minimizes $E[(Y - c)^2]$ is $c = m$.
2. Show that the random variable $f(X)$ that minimizes $E[(Y - f(X))^2 | X]$ is
$$f(X) = E[Y|X].$$
3. Show that the random variable $f(X)$ that minimizes $E[(Y - f(X))^2]$ is also
$$f(X) = E[Y|X].$$

2.4 Conditional Probability and Conditional Expectation

Bayes' theorem

$$\mu(B|A) = \frac{\mu(B)\mu(A|B)}{\mu(A)}$$

Proof

$$\mu(B|A)\mu(A) = \mu(B)\mu(A|B) = \mu(A \cap B)$$

2.4 Conditional Probability and Conditional Expectation

Review of the facebook interview question from Lecture 1

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?

$$P(rain|y,y,y) = \frac{P(y,y,y|rain)P(rain)}{P(y,y,y)}$$

$$= \frac{(2/3)^3 P(rain)}{P(y,y,y)}$$

Can you calculate the denominator? Can you represent it in terms of $P(rain)$?

2.4 Conditional Probability and Conditional Expectation

Exercise Is the conditional probability larger than the prior probability? That is, can you show that

$$\mu(B|A) \geq \mu(B)?$$

This statement implies that collecting data, A, increases the probability of B.

2.4 Conditional Probability and Conditional Expectation

Exercise Is the conditional probability larger than the prior probability? That is, can you show that

$$\mu(B|A) \geq \mu(B)?$$

This statement implies that collecting data, A, increases the probability of B.

Answer: It is not always true. As a counterexample, consider the case $\mu(A) = \mu(B) = 1/2$ and $\mu(A \cap B) = 1/8$. Then $\mu(B|A) = 1/4 < 1/2 = \mu(B)$.

2.4 Conditional Probability and Conditional Expectation

Exercise Is the conditional probability larger than the prior probability? That is, can you show that

$$\mu(B|A) \geq \mu(B)?$$

This statement implies that collecting data, A, increases the probability of B.
Answer: It is not always true. As a counterexample, consider the case $\mu(A) = \mu(B) = 1/2$ and $\mu(A \cap B) = 1/8$. Then $\mu(B|A) = 1/4 < 1/2 = \mu(B)$.
This example shows that collecting data does not alway improve your probability.

2.4 Conditional Probability and Conditional Expectation

Exercise Is the conditional probability larger than the prior probability? That is, can you show that

$$\mu(B|A) \geq \mu(B)?$$

This statement implies that collecting data, A, increases the probability of B.

Answer: It is not always true. As a counterexample, consider the case $\mu(A) = \mu(B) = 1/2$ and $\mu(A \cap B) = 1/8$. Then $\mu(B|A) = 1/4 < 1/2 = \mu(B)$.

This example shows that collecting data does not alway improve your probability.

But wait until the next lecture. There is more to discuss before giving up collecting data.

2.5 Inequalities

Markov's inequality Let X be a non-negative random variable and suppose $E[X]$ exists. For any $t > 0$,

$$\mu(X > t) \leq \frac{E[X]}{t}.$$

2.5 Inequalities

Markov's inequality Let X be a non-negative random variable and suppose $E[X]$ exists. For any $t > 0$,

$$\mu(X > t) \leq \frac{E[X]}{t}.$$

Proof.

$$E[X] = \int_0^\infty xp(x)dx = \int_0^t xp(x)dx + \int_t^\infty xp(x)dx$$

$$\geq \int_t^\infty xp(x)dx \geq t\int_t^\infty p(x)dx = t\mu(X > t).$$

2.5 Inequalities

Chebyshev's inequality Let $m = E[X]$ and $\sigma^2 = Var(X)$. Then,

$$\mu(|X - m| \geq t) \leq \frac{\sigma^2}{t^2} \quad \text{and} \quad \mu(|Z| \geq k) \leq \frac{1}{k^2}$$

where $Z = (X - m)/\sigma$.

2.5 Inequalities

Chebyshev's inequality Let $m = E[X]$ and $\sigma^2 = Var(X)$. Then,

$$\mu(|X - m| \geq t) \leq \frac{\sigma^2}{t^2} \quad \text{and} \quad \mu(|Z| \geq k) \leq \frac{1}{k^2}$$

where $Z = (X - m)/\sigma$.

Proof. Use the Markov's inequality for $Y = |X - m|^2$.

2.5 Inequalities

Exercise Will you consider a coin asymmetric if after 1000 coin tosses the number of heads is equal to 600?

2.6 Types of convergence

Let we have a sequence of random variables,$X_1, X_2, ..., X_n$ and let X is another random variable. Then

- X_n converges to X_n in quadratic mean (or in L_2) if

$$E[(X_n - X)^2] \to 0.$$

- X_n converges to X in probability if for every $\epsilon > 0$

$$\mathbb{P}(|X_n - X| > \epsilon) \to 0$$

- X_n converges to X in distribution if for all t

$$\lim_{n\to\infty} F_n(t) = F(t)$$

  where $F_n(t)$ and $F(t)$ are the distribution functions of X_n and X respectively.

2.6 Types of convergence

Convergence in quadratic mean $\Rightarrow$ Convergence in probability $\Rightarrow$ Convergence in distribution

2.7 Limit Theorems

Let $X_1, X_2, ..., X_n$ are independent, identically distributed random variables with variance σ^2 and mean m.

Q1 What is the mean of $X_1 + X_2 + \cdots + X_n$?

Q2 What is the variance of $X_1 + X_2 + \cdots + X_n$?

2.7 Limit Theorems

Let $X_1, X_2, ..., X_n$ are independent, identically distributed random variables with variance σ^2 and mean m.

Q1 What is the mean of $X_1 + X_2 + \cdots + X_n$?

Q2 What is the variance of $X_1 + X_2 + \cdots + X_n$?

The Law of Large Numbers For

$$\overline{X}_n = \frac{1}{n}\sum_{i}^{n} X_i,$$

converges in probability to the expectation $E[X_i] = m$.

The Central Limit Theorem

Define

$$S_n = \frac{1}{\sqrt{n}}\sum_{i}^{n} X_i.$$

Then S_n converges in distribution to a Gaussian variable with mean m and variance σ^2.

2.7 Limit Theorems

Monte Carlo Integration

$$\int_{[0,1]^d} f(x)dx \approx \frac{1}{n}\sum_{i}^{n} f(x_i)$$

where $\{x_i\}$ is a sample of $[0,1]^d$.
The Central Limit Theorem implies that the Monte Carlo approximation error is of order $\frac{1}{\sqrt{n}}$.

Homework

- Write a code that generates a sample of n values from the standard normal distribution $N(0,1)$. n is an input parameter of the code.
- Draw a histogram of the sample.
- Draw the Gaussian fit to the sample statistics. That is, draw the Gaussian density with the same mean and variance of the sample.
- Draw a histogram of $y_i = e_i^x$ where x_i is a sample from the standard normal distribution.
- Write a code that draws a sample of n values of the uniform distribution on $[0,1]$. n is an input parameter of the code.
- Use a transformation of random variables to generate samples from the Cauchy density $p(x) = \frac{1}{\pi(1+x^2)}$.
- Draw a histogram of the sample.
- Calculate the mean. Plot the mean as a function of n.