COMPSCI 240: Reasoning Under Uncertainty

Nic Herndon and Andrew Lan

University of Massachusetts at Amherst

Spring 2019

Logistics

- We'll try to get grading done this week
- Homework 2 grading will have to wait
- How hard was the exam?
- No homeworks, only quizzes until spring break

Lecture 13: Continuous Random Variables

Cont. Multiple Random Variables

- Given random variables $X_1, X_2, \ldots, X_N$ and a function $f : \mathbb{R} \times \mathbb{R} \times \ldots \times \mathbb{R} \to \mathbb{R}$,

$$Z = f(X_1, X_2, \ldots, X_N)$$

  is a new random variable.

- Linearity of Expectation: If $Z = \sum_{i=1}^{N} X_i$,

$$E[Z] = E\left[\sum_{i=1}^{N} X_i\right] = \sum_{i=1}^{N} E[X_i]$$

- Independence: If $X_1, \ldots, X_N$ are independent,

$$P_{X_1,\cdots,X_N}(x_i, \cdots, x_N) = \prod_{i=1}^{N} P_{X_i}(x_i) \ .$$

- Linearity of Variance: If $Z = \sum_{i=1}^{N} X_i$ and all X_i are independent,

$$var[Z] = var\left(\sum_{i=1}^{N} X_i\right) = \sum_{i=1}^{N} var[X_i]$$

Example 1

Toss a fair six-sided dice. Let X be that you see "1" $(X = 1)$ or not $(X = 0)$, and let Y be that you see a "6" $(Y = 1)$ or not $(Y = 0)$. Are X and Y independent:

A: Yes

B: No

C: Can't tell from the information given.

Answer is **B** since $P(X = 1, Y = 1) = 0 \neq P(X = 1)P(Y = 1)$. For example, knowing that X has occurred can carry important information regarding the occurrence of Y.

Example 2

Toss 12 fair six-sided dice. Let X be the number of "1"s and let Y be the number of "6"s. What is the expected value of X?

A: 0

B: 1

C: 2

D: 3

E: 6

Answer is **C** because X is a binomial random variable with $n = 12$ and $p = 1/6$. Thus,

$$E[X] = np = 2.$$

Example 3

- Toss 12 fair six-sided dice. Let X be the number of "1"s and let Y be the number of "6"s. What is the expected value of $X + Y$?

  A: 0
  B: 2
  C: 4
  D: 6
  E: 12

  Answer is **C** because $E[X + Y] = E[X] + E[Y] = 2 + 2 = 4$.

- Would $Var[X + Y] = Var[X] + Var[Y]$ hold?
  Answer is **NO** because X and Y are not independent.

A Slightly Challenging Joint PMF Example

Bob is a new TA for one coursework. He has a probability of $1/4$ to answer one question incorrectly independent of other questions. In a class, Bob can be asked a maximum of 3 questions. The probability that Bob will be asked 1, 2, and 3 questions will be $1/3$, respectively; Bob is guaranteed to be asked at least one question in that class. Let X be the number of questions Bob is asked and Y be the questions he answers incorrectly.
Construct the joint PMF table.

P(X,Y)				
X\Y	0	1	2	3
1	1/4	1/12	0	0
2	3/16	1/8	1/48	0
3	9/64	9/64	3/64	1/192

A Slightly Challenging Joint PMF Example

Bob is a new TA for one coursework. He has a probability of 1/4 to answer one question incorrectly independent of other questions. In a class, Bob can be asked a maximum of 3 questions. The probability that Bob will be asked 1, 2, and 3 questions will be 1/3, respectively; Bob is guaranteed to be asked at least one question in that class. Let X be the number of questions Bob is asked and Y be the questions he answers incorrectly.

What is the probability that Bob was asked 2 questions given that he has answered 1 question incorrectly?

P(X,Y)				
X\Y	0	1	2	3
1	1/4	1/12	0	0
2	3/16	1/8	1/48	0
3	9/64	9/64	3/64	1/192

$$P(X=2, Y=1) = \frac{1}{8}$$

$$P(X=2|Y=1) = \frac{P(X=2, Y=1)}{P(Y=1)} = \frac{1/8}{1/12 + 1/8 + 9/64}$$

Probability Laws

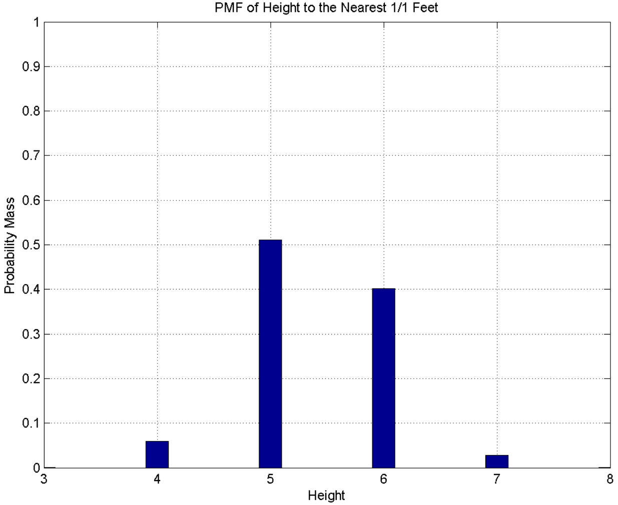
- Probability law: assigns a probability $P(A)$ to any event $A \subseteq \Omega$ encoding our knowledge or beliefs about the collective "likelihood" of the elements of event A; satisfies 3 axioms:

- Nonnegativity: $P(A) \geq 0$ for every $A \subseteq \Omega$
- Additivity: $P(A \cup B) = P(A) + P(B)$ if A and B are disjoint
- Normalization: $P(\Omega) = 1$

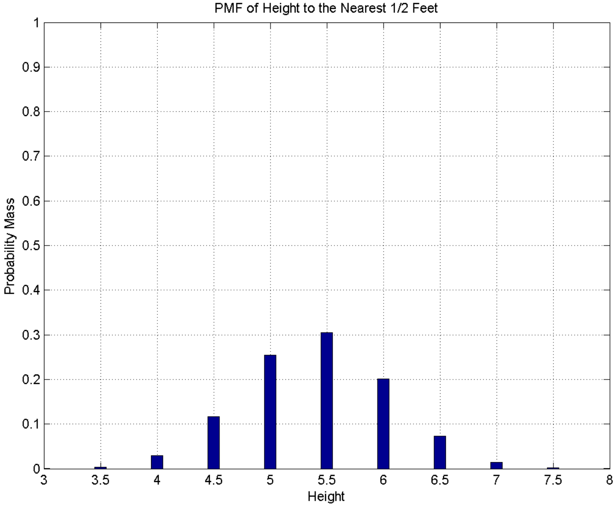
Continuous Random Variables

- There are many random variables that are much more naturally thought of as taking continuous values than a finite or countable number of values (ex: height, weight, distance, time, speed, cost, etc...).
- **Example:** Suppose we measure the height of a randomly selected person to the nearest foot. This gives a discrete random variable with a probability mass function.
- **Question:** What happens to the probability mass function of height if we measure it in smaller and smaller units?

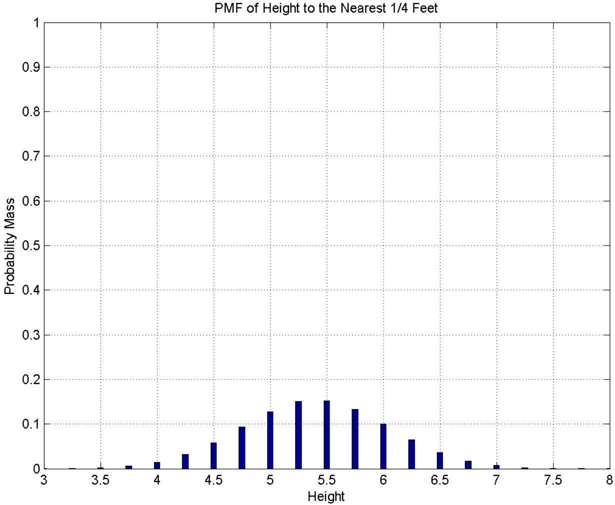
Example: Height - Nearest Foot



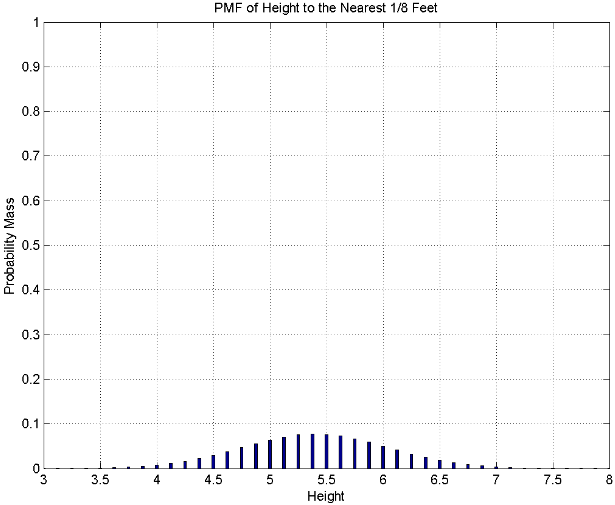
Example: Height - Nearest 1/2 Foot



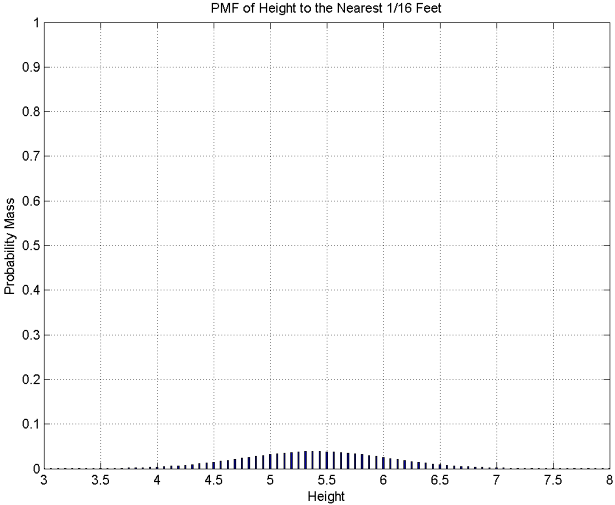
Example: Height - Nearest 1/4 Foot



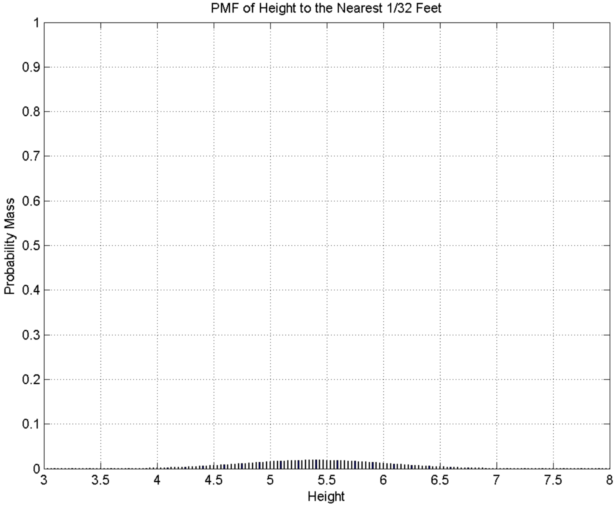
Example: Height - Nearest 1/8 Foot



Example: Height - Nearest 1/16 Foot



Example: Height - Nearest 1/32 Foot



Example: Height

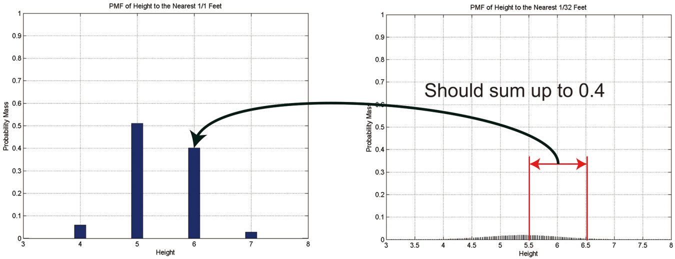
- **Question:** What happens to the probability mass function if we measure the height of people in smaller and smaller units?
- **Answer:** As we measure the height of people in smaller and smaller units, the number of values in the range of the random variable becomes uncountably infinite and the probability mass assigned to any particular value goes to zero!

Probability of Intervals

- However, the probability mass $P(a < X < b)$ associated with an **interval** $[a, b]$ of $\mathcal{X}$ of length greater than zero can be non-zero.

Probability of Intervals

- **Example:** In the height example, the probability that the height of a person was between 5.5 and 6.5 feet was about 0.4. In other words $P(5.5 < X < 6.5) = 0.4$.



- In general, if $A \subseteq \mathcal{X}$ is any subset of the range of X that has non-zero length, then $P(X \in A)$ can be non-zero.

Continuous Random Variable

- We see that continuous random variable can provide more fine-grained probability profiles of a random variable.
- Is that it? In the height example, in the real world, we don't need to have high resolutions of $1/32$ feet.
- Most of real-world applications have discrete measurements (e.g., height, speed, etc.)
- Continuous random variables allow the use of powerful tools from calculus and often admit an insightful analysis that would not be possible under a discrete model.

Probability Density

- The standard way to construct probability laws for continuous random variables is using a **probability density function**.
- The only restrictions on the density function are that
  - **Non-negativity**: $f_X(x) \geq 0$ for all $X \subseteq \mathcal{X}$
  - **Normalization**: $\int_{\mathcal{X}} f_X(x)dx = \int_{-\infty}^{\infty} f_X(x)dx = 1$
- If we have a random variable X with probability density function $f_X(x)$ then the probability of any set $A \subseteq \mathcal{X}$ is given by the integral of the density over A:

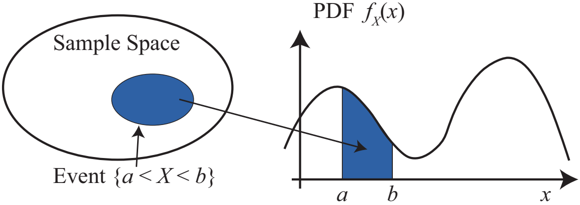
$$P(X \in A) = \int_A f_X(x)dx$$

Probability Density

- In the simplest case $A = [a, b]$ is a single interval and this definition reduces to a definite integral:

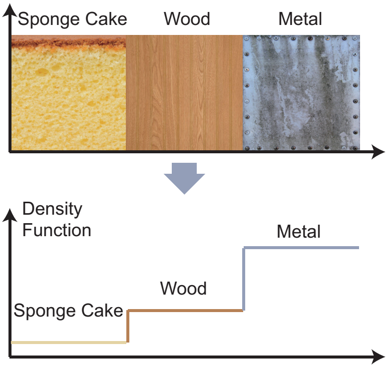
$$P(a < X < b) = \int_a^b f_X(x)dx$$

- Intuitively, the **probability mass** of an interval $[a, b]$ is $P(a < X < b)$.



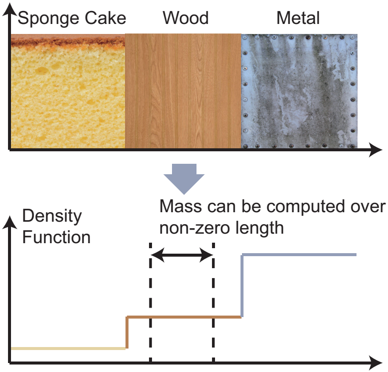
Probability Density - Conceptual Visualization

- Image you have a thin panels of 1) sponge cake, 2) wood, and 3) metal.



Probability Density - Conceptual Visualization

- Image you have a thin panels of 1) sponge cake, 2) wood, and 3) metal.



- This is why we name it PDF and PMF.

Probability Density - Additivity

- When these two requirements (non-negativity and normalization) are met, defining probabilities of sets by integrating the density function over the set will automatically satisfy the additivity axiom.
- For example, if $[a, c]$ is any interval and $a < b < c$ then:

$$\begin{aligned} P(a < X < c) &= \int_a^c f_X(x)dx \\ &= \int_a^b f_X(x)dx + \int_b^c f_X(x)dx \\ &= P(a < X < b) + P(b < X < c) \end{aligned}$$

- However, if $[a, d]$ is any interval and $a < b < c < d$ then:

$$\begin{aligned} P(a < X < d) &= \int_a^d f_X(x)dx \\ &\neq \int_a^c f_X(x)dx + \int_b^d f_X(x)dx \\ &\neq P(a < X < c) + P(b < X < d) \end{aligned}$$

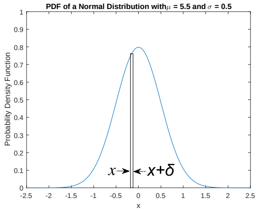
Probability Density

- For a single value a, we have

$$P(X = a) = \int_a^a f_X(x)dx = 0$$

- More precisely,

$$\begin{aligned} P(X = a) &= P(a < X < a + \delta) \text{ where } \delta \approx 0 \\ &= \int_x^{x+\delta} f_X(t)dt \approx f_X(x) \cdot \delta \\ &= 0 \end{aligned}$$



- For this reason,

$$P(a \le X \le b) = P(a < X < b) = P(a \le X < b) = P(a < X \le b)$$

Expectation and Variance

- The expected value or mean of a continuous random variable X can be computed in a similar manner as a discrete random variable:

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

- The expected value of a function of a random variable $g(X)$ can be computed as

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

- The variance of X can be computed as

$$\begin{aligned} var(X) &= E\left[(x - E(x))^2\right] \\ &= \int_{-\infty}^{\infty} [x - E(x)]^2 f_X(x) dx \\ &= E(X^2) - E(X)^2 \end{aligned}$$