IV. Random Variables

Instructor: **Yanlin Qi**

Institute of Transportation Studies
Department of Statistics
University of California, Davis

Today

- Random variables (discrete, continuous)
- Probability Mass Functions (PMF)
- Probability Density Functions (PDF)
- Cumulative Distribution Functions (CDF)
- Review Examples

Random Variables

Random Variables

- Because:
  - The particular outcome of an experiment is not known in advance
  - The resulting value of a variable is not known in advance
- A Random Variable (RV) is a function that assigns a real number to **each outcome** in the sample space of a random experiment

Notation

- A RV is denoted by an uppercase letter such as X
  - Examples:
    - C = A coin toss
    - S = Strength of a material
    - H = Height of students in a class
- After an experiment is conducted, the measured value of the RV is denoted by a lowercase letter such as x
  - Examples:
    - c = The outcome of tossing a coin (head or tail)
    - s = Experimental strength of a material specimen
    - h = Measure height of a student in a class

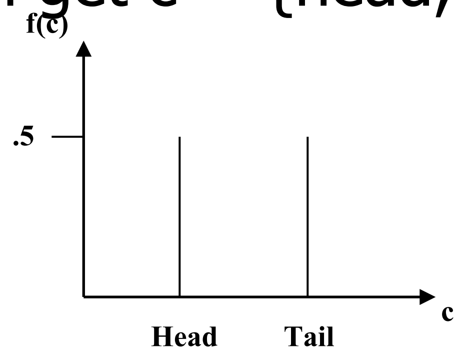
Discrete and Continuous

- A discrete RV has a finite (or countably infinite) range
  - Examples:
    - Number of scratches on a surface
    - Proportion of defective parts among 1000 tested
    - Number of transmitted bits received in error, …
- A continuous RV has an interval (either finite or infinite) of real numbers for its range
  - Examples:
    - Material strength
    - Electrical current
    - Length
    - Pressure
    - Temperature, …

Distribution Functions

Distribution Functions

- Distribution functions are commonly used in engineering
- They are a way to assign probabilities to each possible value the variable can take
- The distribution function of a RV is a description of the probabilities associated with its possible values. It can be represented with a function (f)
- Example 1:
  - When you toss a coin (C), you can get c = {head, tail}
    - Each with probability of 0.5



Discrete vs Continuous

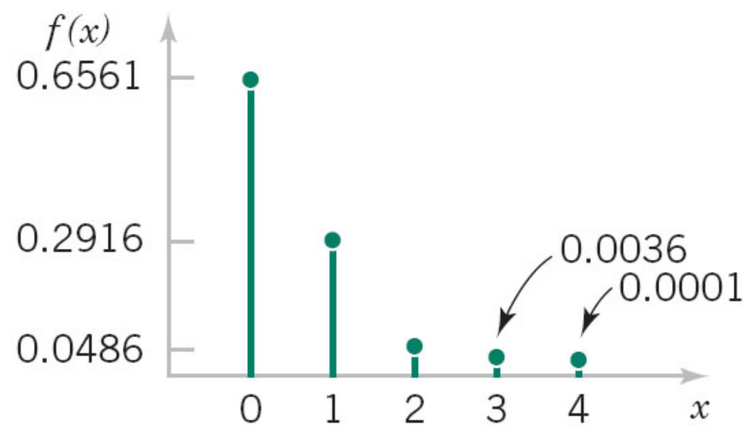
Type of Variable	Distribution Function
Discrete	Probability Mass Function (PMF)
Continuous	Probability Density Function (PDF)

Probability Mass Functions (PMFs)

- PMFs $f(x_i)$ can be used to describe the probability distribution of a discrete RV X
- Example 2:

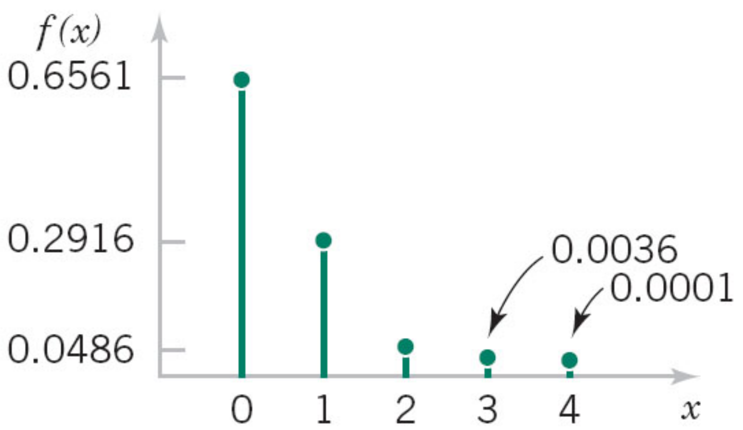
The number of points you can get from an exercise ranges from 0 to 4. However, you can not get intermediate points. That is you can only get 0, 1, 2, 3, and 4. Based on historic data, a grading model that estimates de probabilities of getting each score was developed. The estimated probabilities are:

$P(X=0)=f(0) = 0.6561$
$P(X=1)=0.2916$
$P(X=2)=0.0486$
$P(X=3)=0.0036$
$P(X=4)=0.0001$



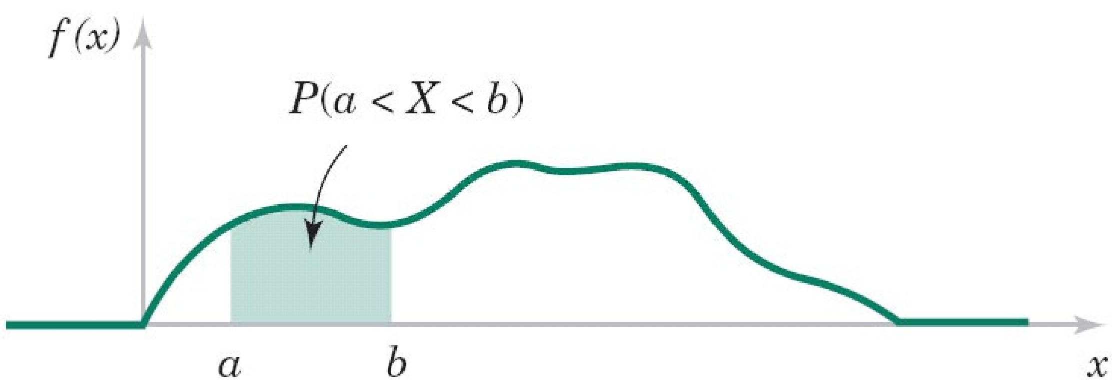
Probability Mass Functions (PMFs)

- For a discrete RV X, a PMF is a function such that
  - $f(x_i) \geq 0$
  - $\sum_{i=1}^{n} f(x_i) = 1$
  - $P(X = x_i) = f(x_i)$
- The probability that X is between a and b is determined as the summation of $f(x_i)$ from a to b
  - $(a$ to $b)$ is the range of the RV X



Probability Density Functions (PDFs)

- A PDF $f(x)$ is a function that describes the probability distribution of a continuous RV X
  - *The likelihood of a continuous RV taking on a specific value*
- If an interval is likely to contain a value for X, its probability is larger with larger values for $f(x)$
- Note: f (x) itself is not a probability; it is a density

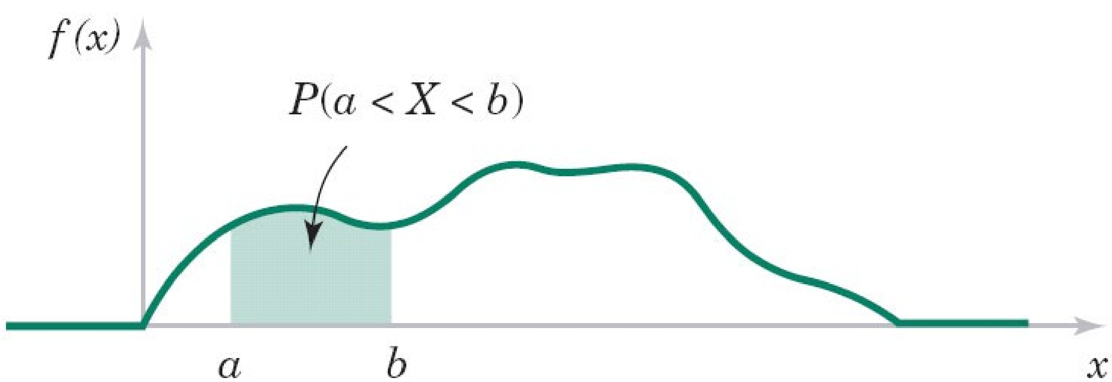


Probability Density Function (PDF)

- For a continuous RV X, a PDF is a function such that
  - $f(x) \geq 0$
  - $\int_{-\infty}^{\infty} f(x)dx = 1$
  - $P(a \leq X \leq b) = \int_{a}^{b} f(x)dx =$ area under $f(x)$ from a to b

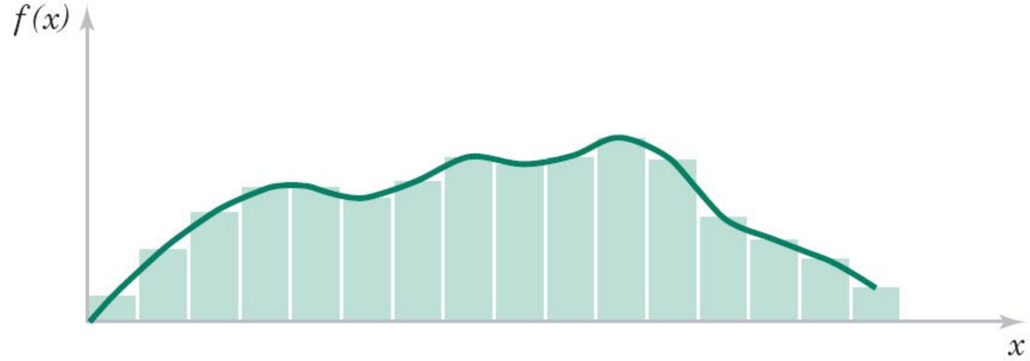
  Valid probability distribution

- The probability that X is between a and b is determined as the integral of $f(x)$ from a to b



Approximating PDFs

- A histogram is an approximation to a PDF
- For each interval of the histogram, the area of the bar equals the relative frequency of the measurements in the interval
  - This is an estimate of the probability that a measurement falls in the interval
- Therefore: $P(X = x) = 0$ (Think about why?)
  - Also:
  - $P(x_1 \le X \le x_2) = P(x_1 < X \le x_2) = P(x_1 \le X < x_2) = P(x_1 < X < x_2)$



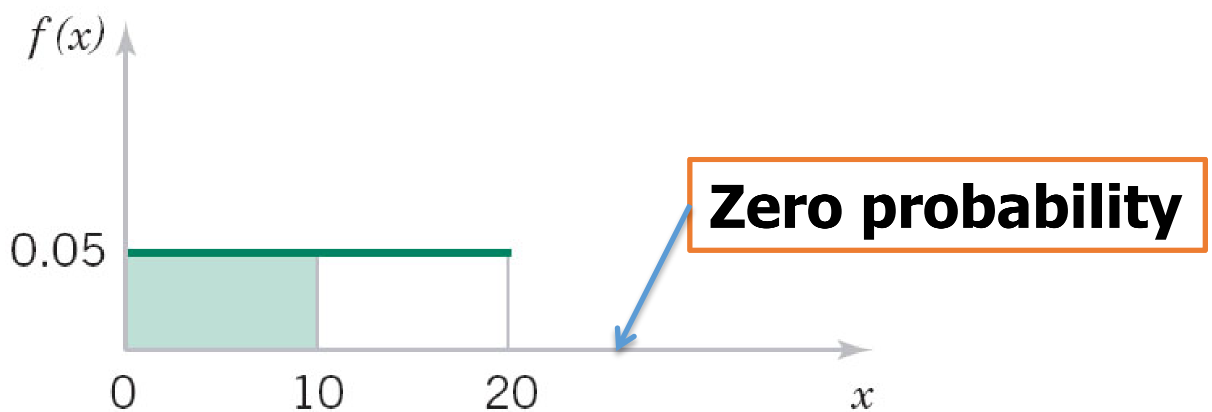
Example 3

- Let X = concrete strength, in the range $[0, 20\ \text{MPa}]$
- Assume a PDF given as:

  $f(x) = 0.05$ for $0 \le x \le 20$

- What is the probability that the strength of a tested concrete is less than 10 MPa?

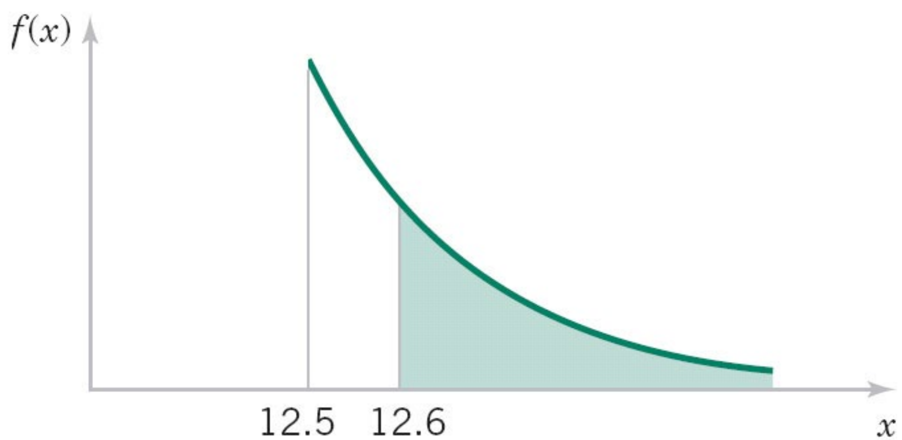
$$P(X < 10) = \int_0^{10} f(x)dx = \int_0^{10} 0.05dx = 0.5$$



Example 4

- Let X = the length of a construction steel bar
- Production target length is 12.5 m. However, the length of the steel bars is affected by disturbances in the extrusion process
- From available data, the PDF can be described as

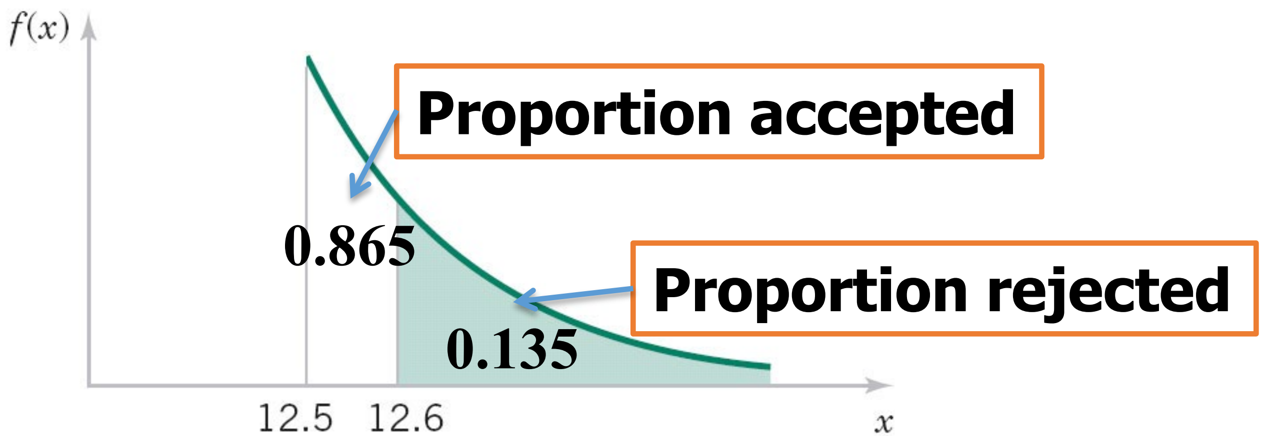
$$f(x) = 20e^{-20(x-12.5)}, \ x \geq 12.5$$



Example 4

- If a piece with length of 12.6 m is rejected, what is the proportion of pieces rejected?

$$P(X > 12.60) = \int_{12.6}^{\infty} f(x)dx = \int_{12.6}^{\infty} 20e^{-20(x-12.5)}dx = -e^{-20(x-12.5)}\Big|_{12.6}^{\infty} = 0.135$$



Cumulative Distributions

Cumulative Distribution Function (CDF)

- Alternative method to describe RVs
- The CDF of a discrete RV X is

$$F(x) = P(X \le x) = \sum_{x \le x_i} f(x_i)$$

- The CDF of a continuous RV X is

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(u)du$$

- Note that a CDF provides probabilities

Properties of a CDF for a discrete RV

- For a discrete RV X, $F(x)$ satisfies the following properties:

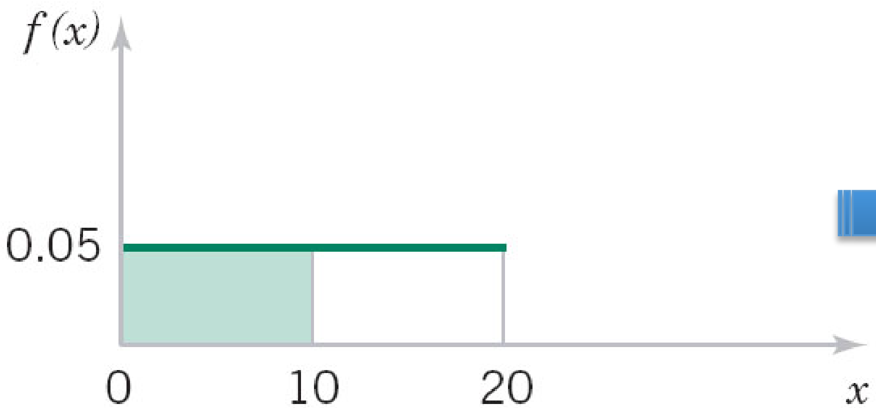
1. $$F(x) = P(X \le x) = \sum_{x \le x_i} f(x_i)$$

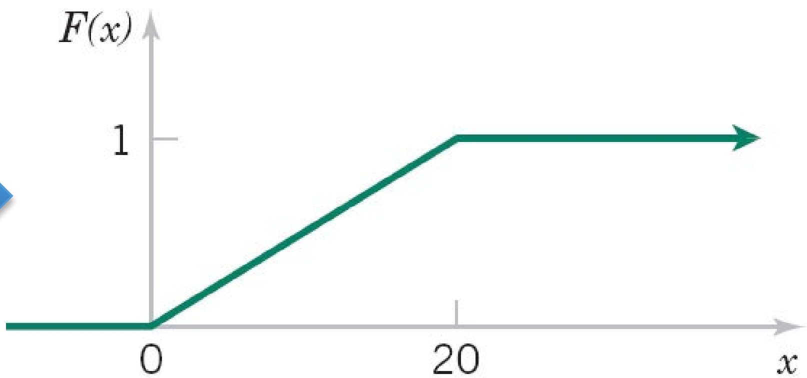
2. $$0 \le F(x) \le 1$$

3. $$\text{if } x \le y, \text{ then } F(x) \le F(y)$$

Example 5

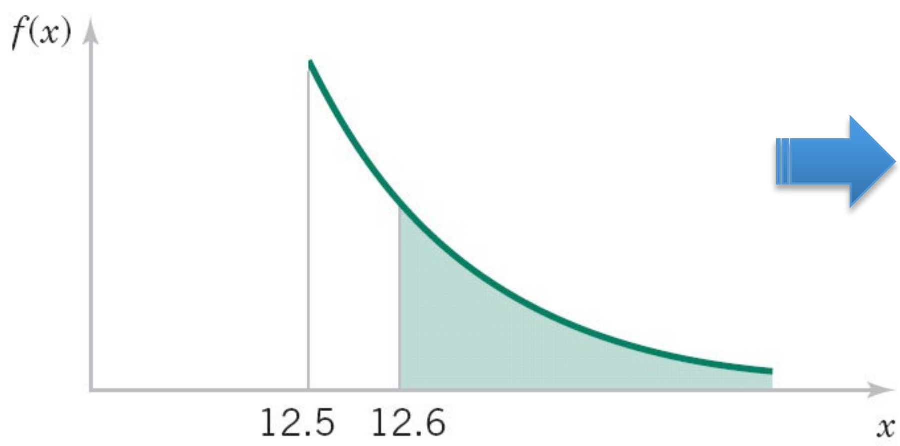
- For the concrete strength test example, we have:
  - for $x < 0 \quad F(x) = 0$
  - for $0 \leq x < 20 \quad F(x) = \int_0^x f(u)du = 0.05x$
  - for $20 \leq x \quad F(x) = \int_{20}^x f(u)du = 1$

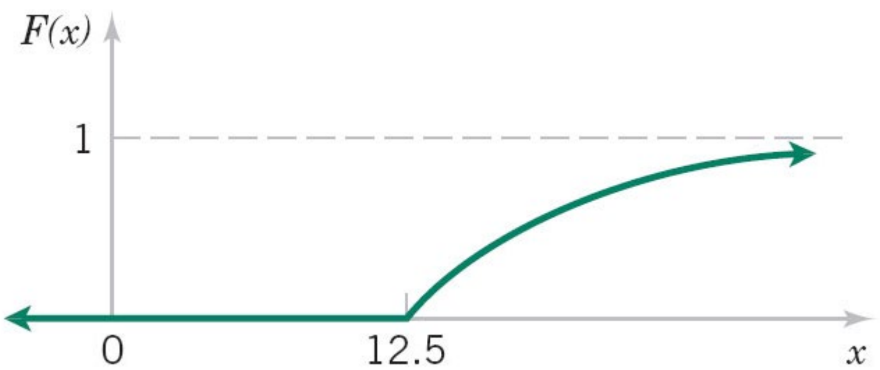




Example 6

- For the construction steel bar example we have:
  - for $x < 12.5 \quad F(x) = 0$
  - $\text{for } 12.5 \le x \quad F(x) = \int_{12.5}^{x} 20e^{-20(u-12.5)}du = 1 - e^{-20(x-12.5)}$





Differential Relationship between PDF and CDF

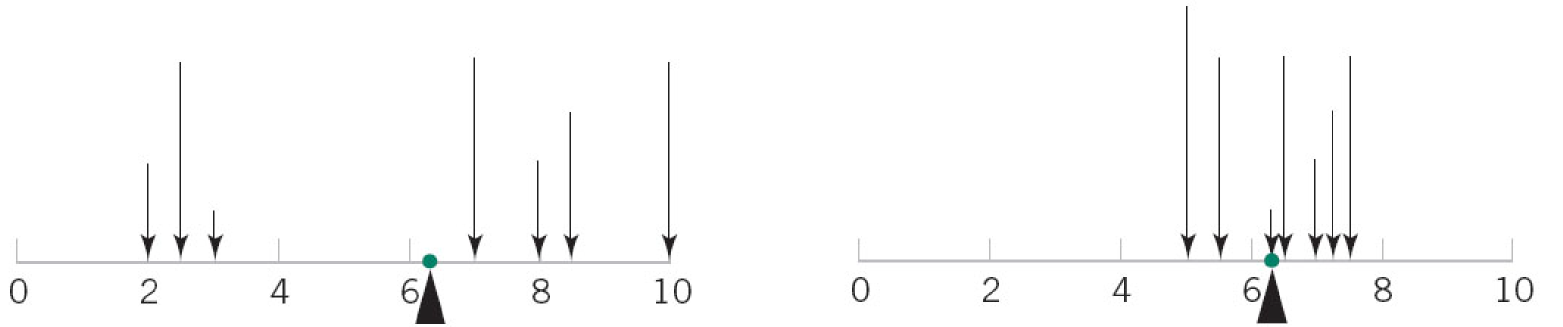
- The PDF of a continuous RV can be obtained from the CDF by differentiating

$$f(x) = \frac{dF(x)}{dx}$$

Mean and Variance

Mean and Variance of a RV

- Two numbers are used to summarize a Probability Mass/Density Function P(M/D)F for a RV X
  - The mean or expected value: measure of the center or middle of a P(M/D)F
  - The variance: measure of the dispersion, or variability
  - They do not identify a P(M/D)F
    - Different P(M/D)Fs can have the same mean or variance
    - Example 7:
      - If a P(M/D)F is viewed as a loading on a beam, the mean is the balance point



Mean and Variance of a Discrete RV

- Suppose X is a discrete RV with PMF $f(x_i)$
- The mean or expected value of X
  - Denoted as μ or $E(X)$, is

$$\mu = E(X) = \sum_{i=1}^{n} x_i f(x_i)$$

- The variance of X
  - Denoted as σ^2 or $V(X)$, is

$$\sigma^2 = V(X) = E(X-\mu)^2 = \sum_{1=1}^{n} (x_i - \mu)^2 f(x_i) = \sum_{i=1}^{n} x_i^2 f(x_i) - \mu^2$$

- The standard deviation of X is

$$\sigma = \sqrt{\sigma^2}$$

Mean and Variance of a Continuous RV

- Suppose X is a continuous RV with PDF $f(x)$
- The mean or expected value of X
  - Denoted as μ or $E(X)$, is

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

- The variance of X
  - Denoted as σ^2 or $V(X)$, is

$$\sigma^2 = V(X) = E(X-\mu)^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx = \int_{-\infty}^{\infty} x^2 f(x) dx - \mu^2$$

- The standard deviation of X is

$$\sigma = \sqrt{\sigma^2}$$

Equality of the Formulas for Variance

- Properties of integrals and the definition of mean are used
  - For a discrete RV
  - For a continuous RV

Example 8

- For the concrete strength, the mean of X with $f(x) = 0.05$ is

$$E(X) = \int_0^{20} x\, f(x)\, dx = \left.\frac{0.05x^2}{2}\right|_0^{20} = 10$$

- The variance of X is

$$V(X) = \int_0^{20} (x-10)^2 f(x)\, dx = \left.\frac{0.05(x-10)^3}{3}\right|_0^{20} = 33.3$$

Expected Value of a Function of a RV

- It is defined in a straightforward manner as

$$E[h(X)] = \sum_{i=1}^{n} h(x_i) f(x_i) \qquad E[h(X)] = \int_{-\infty}^{\infty} h(x) f(x) dx$$

- Example 9:
  - In the example of the concrete strength, what is the expected value of the squared strength?
  - $h(X) = X^2$, therefore

$$E[h(X)] = \int_{-\infty}^{\infty} x^2 f(x) dx = \int_{-\infty}^{\infty} 0.05 x^2 f(x) dx = \left. \frac{0.05 x^3}{3} \right|_0^{20} = 133.3$$

  - Note that the expected value of the square does not equal the square of the expected value: $E(X^2) \neq [E(X)]^2$

Summary – Part I

- Discrete and Continuous Random Variables
- Probability Distributions
  - Probability Mass Function
  - Probability Density Function
  - Cumulative Distribution Function
- Mean and Variance of Random Variables

Review Examples

Examples 6

- Determine the range of the random variable:
  - a. The random variables is the number of failed solder connections on a printer circuit board with 1000 connections
  - b. An order for a car can select the base model or add any of 15 options. The random variable is the number of options selected in the order.
  - c. A batch of 500 machined parts contains 10 that do not conform to customer requirements. Parts are selected successively, without replacement, until a nonconforming part is obtained. The random variable is the number of parts selected.

Examples 7

- The sample space of random experiment is {a, b, c, d, e, f}, and each outcome is equally likely. A random variable is defined as:

outcome	a	b	c	d	e	f
x	0	0	1.5	1.5	2	3

- Determine the PMF of X. Use the PMF to determine:
  a. P(X = 1.5)
  b. P(0.5 < X < 2.7)
  c. P(X>3)
  d. P(0<= X < 2)
  e. P (X=0 or X =2)
- Determine the Cumulative Distribution Function of X

Examples 8

- In a semiconductor manufacturing process, three pieces from a lot are tested. Each piece is classified as pass or fail. Assume that the probability that a piece passes the test is 0.8 and that each piece is independent.
  - Determine the PMF of the number of pieces from the lot that pass the test.
  - Answer:
    - X = number of wafers that pass

Examples 9

- Trees are subjected to different levels of carbon dioxide atmosphere with 6% of the tree in a minimal growth condition at 350 parts per million (ppm), 10% at 450 ppm (slow growth), 47% at 550 ppm (moderate growth), and 37% at 650 ppm (rapid growth).
- What are the mean and standard deviation of the carbon dioxide atmosphere (in ppm) for these trees in ppm?

Examples 10

- Suppose that $f(x) = e^{(-x)}$ for $0<x$. Determine:
  a. $P(1<X)$
  b. $P(1<X<2.5)$
  c. $P(X=3)$
  d. $P(X<4)$
  e. $P(3<= X)$
  f. Determine x such that $P(x<X)=.1$
  g. Determine x such that $P(X<=x)=.1$
- Determine the cumulative distribute function of the distribution

Examples 11

- Suppose that $f(x) = 3(8x-x^2)/256$ for $0<x<8$. Determine:
  a. $P(X<2)$
  b. $P(X<9)$
  c. $P(2<X<4)$
  d. $P(X>6)$
- Determine the cumulative distribute function of the distribution

Examples 12

- Suppose the PDF of the length of a computer cable is $f(x) = 0.1$ from 1200 to 1210 millimeters.
  a. Determine the mean and standard deviation of the cable length
  b. If the length specifications are $1995 < x < 1205$ millimeters, what proportion of cables are within specification?

Next Class

- Continue with Probability Distributions Functions
  - Uniform, Binomial and Poisson Distributions

S. (3-5, 4-5), 3-6, and 3-9

  - Start with Normal (N) (Section 4-6)
  - Relationship between Binomial and Poisson with N (S. 4-7)
  - Related to binomial:
    - Geometric, Negative Bin., Hypergeometric (S. 3-7, 3-8)