Random Variables

Keegan Korthauer
Department of Statistics
UW Madison

RANDOM VARIABLES

Discrete random variables

Continuous random variables

Distribution functions

Mean and variance

Types of Random Variables

- Discrete
  - Gap between assigned values
  - Finitely or infinitely many countable values
- Continuous
  - Assigned values form continuous intervals
  - Infinitely many uncountable values

Examples of Random Variables

- The number of imperfections in a computer chip (Discrete)
- The highest temperature in Madison next year (Continuous)
- The sum of two rolls of a die (Discrete)
- The lifetime of a light bulb (Continuous)

Random Variable

- A **random variable** is an **assignment of numerical values to outcomes** in a sample space
- Resistor example – draw from 2 boxes each with 3 incorrectly labeled resistors (box 1 actual: 9,10,11; box 2 actual: 19,20,21)
  - Random Variable X = Total resistance when connecting the two

Outcome	X	Probability
(9, 19)	28	1/9
(9, 20)	29	1/9
(9, 21)	30	1/9
(10, 19)	29	1/9
(10, 20)	30	1/9
(10, 21)	31	1/9
(11, 19)	30	1/9
(11, 20)	31	1/9
(11, 21)	32	1/9

Simplify

x	$P(X = x)$
28	1/9
29	2/9
30	3/9
31	2/9
32	1/9

Discrete R.V. - Example

- Surface imperfections on computer chips

Number of Imperfections (Y)	0	1	2	3	4	5
Probability	0.09	0.22	0.26	0.20	0.12	0.11

- Y = number of imperfections in a randomly chosen chip
- What are the possible values for Y?
- What is P(Y>3)?

Discrete R.V. - Example

- Surface imperfections on computer chips

Number of Imperfections (Y)	0	1	2	3	4	5
Probability	0.09	0.22	0.26	0.20	0.12	0.11

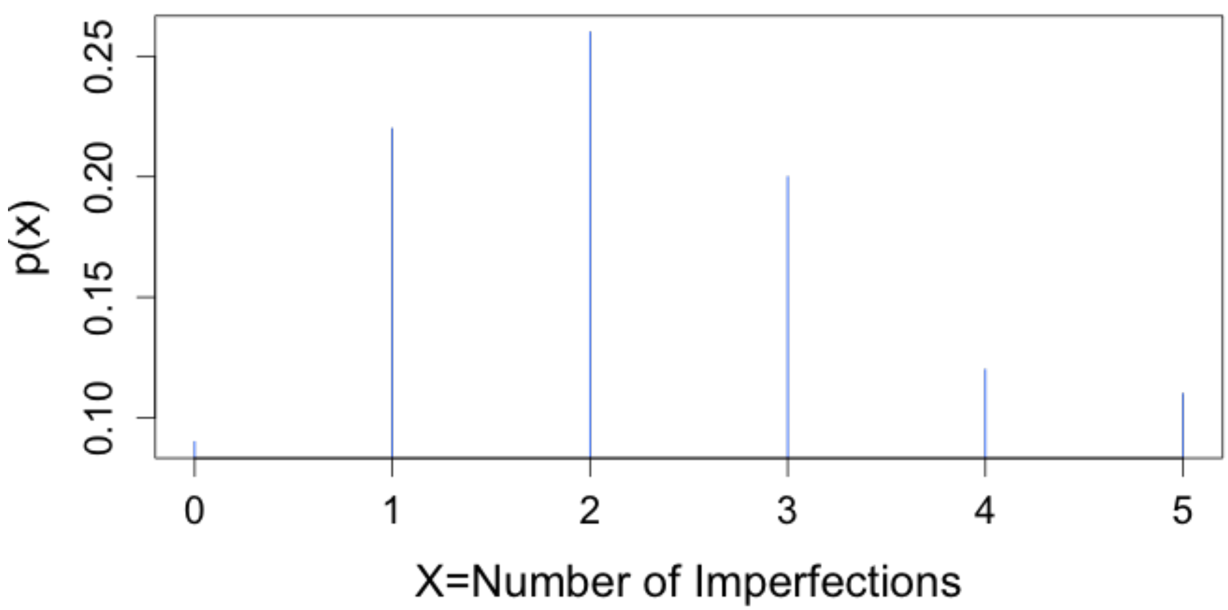
- Y = number of imperfections in a randomly chosen chip
- What are the possible values for Y? {0,1,2,3,4,5}
- What is P(Y>3)? $P(Y>3) = P(Y=4) + P(Y=5)$

  $= 0.12 + 0.11 = 0.23$

Probability Mass Function (Discrete)

Let X be a discrete random variable. The probability mass function (PMF) of X is

$$p(x) = P(X = x)$$



Property of the PMF

The sum of all possible values of p(x) is one:

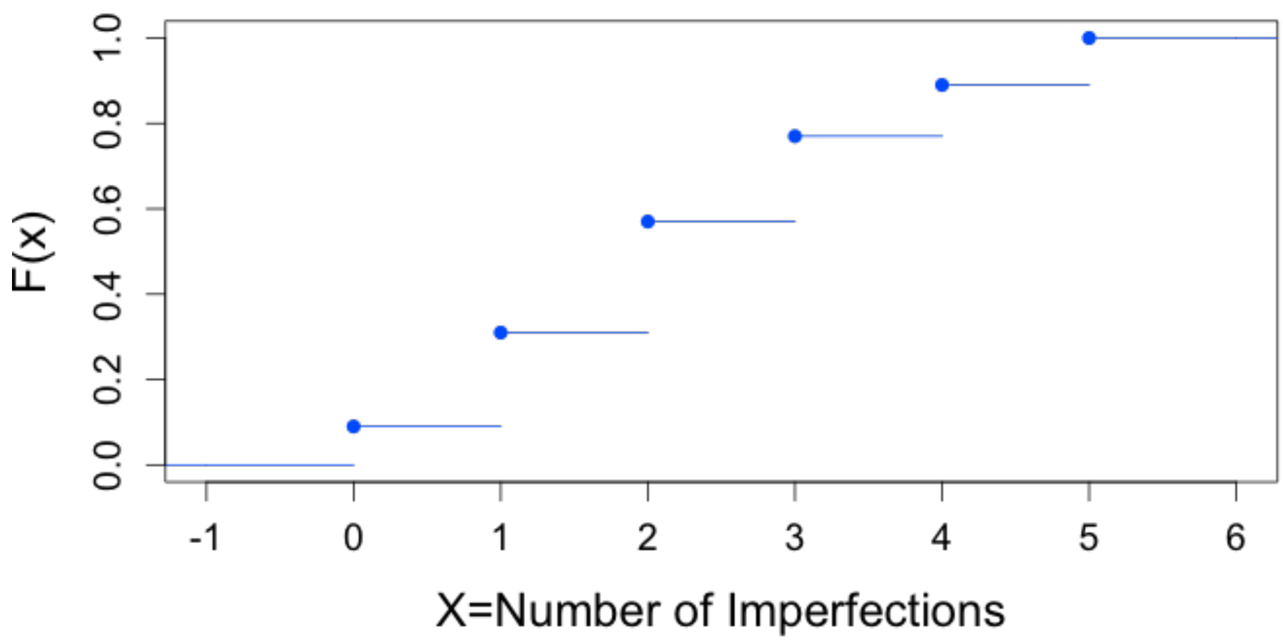
$$\sum_x p(x) = \sum_x P(X = x) = 1$$

where the sum is over all possible values of X

Cumulative Distribution Function (Discrete)

Let X be a discrete random variable. The cumulative distribution function (CDF) of X is

$$F(x) = P(X \leq x) = \sum_{t \leq x} P(X = t)$$



Example – Finding the PMF/CDF

Let Y be the discrete RV representing the sum of numbers on two die. Enumerate the sample space and specify the PMF and CDF.

Population Mean (Discrete)

- Let X be a discrete random variable. The **mean** of X is given by

$$\mu_X = \sum_x xP(X = x)$$

- Sum is over all possible values of X
- Also called the **expectation** or **expected value**, can be denoted as **E(X)** or **μ**

Population Variance (Discrete)

- Let X be a discrete random variable. The **variance** of X is given by

$$\sigma_X^2 = \sum_x (x - \mu_X)^2 P(X = x)$$
$$= \sum_x x^2 P(X = x) - \mu_X^2$$

- May also be denoted by **Var(X)**, **V(X),** or simply $\boldsymbol{\sigma^2}$
- **Standard deviation** is the square root of the variance, denoted **σ** or **sd(X)**

Example – Finding Mean and Variance

A certain community is surveyed for how many cars it contains. Let X represent the number of cars per household and assume that X has the following probability mass function:

Number of Cars (X)	0	1	2	3	4
Probability	0.10	0.25	0.55	0.09	0.01

Find the mean and variance of X

Example – Finding Mean and Variance

A certain community is surveyed for how many cars it contains. Let X represent the number of cars per household and assume that X has the following probability mass function:

Number of Cars (X)	0	1	2	3	4
Probability	0.10	0.25	0.55	0.09	0.01

Find the mean and variance of X

$E(X) = 1.66$ cars
$V(X) = 0.6644 \text{ cars}^2$

Exercise 2.4.12

Suppose we have a collection of components to test (success = S and failure = F), each with failure probability of 0.2. Let X represent the number of successes among a sample of three components.

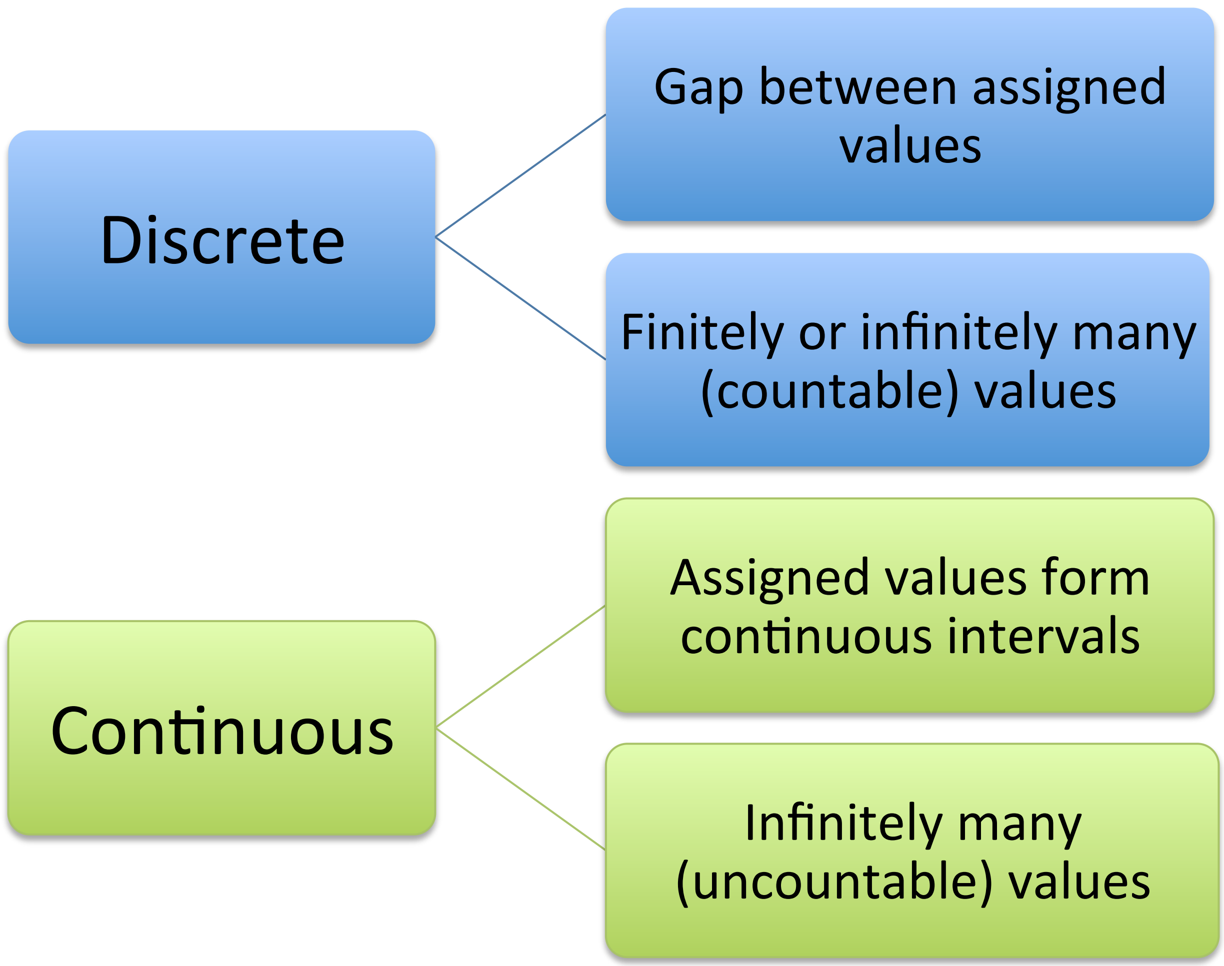
1. What are the possible values for X?
2. Find P(X=3)
3. The event that the first component fails and the next two succeed is denoted FSS. Find P(FSS)
4. Find P(SFS) and P(SSF)
5. Find P(X=0), P(X=1) and P(X=2)
6. Find E(X) and V(X)

Exercise 2.4.12

Suppose we have a collection of components to test (success = S and failure = F), each with failure probability of 0.2. Let X represent the number of successes among a sample of three components.

1. What are the possible values for X? 0,1,2,3
2. Find P(X=3) = 0.512
3. The event that the first component fails and the next two succeed is denoted FSS. Find P(FSS) = 0.128
4. Find P(SFS) and P(SSF) = 0.128
5. Find P(X=0) = 0.008, P(X=1) = 0.096 and P(X=2) = 0.384
6. Find E(X) = 2.4 and V(X) = 0.48

Types of Random Variables



Probability Histogram (Discrete)

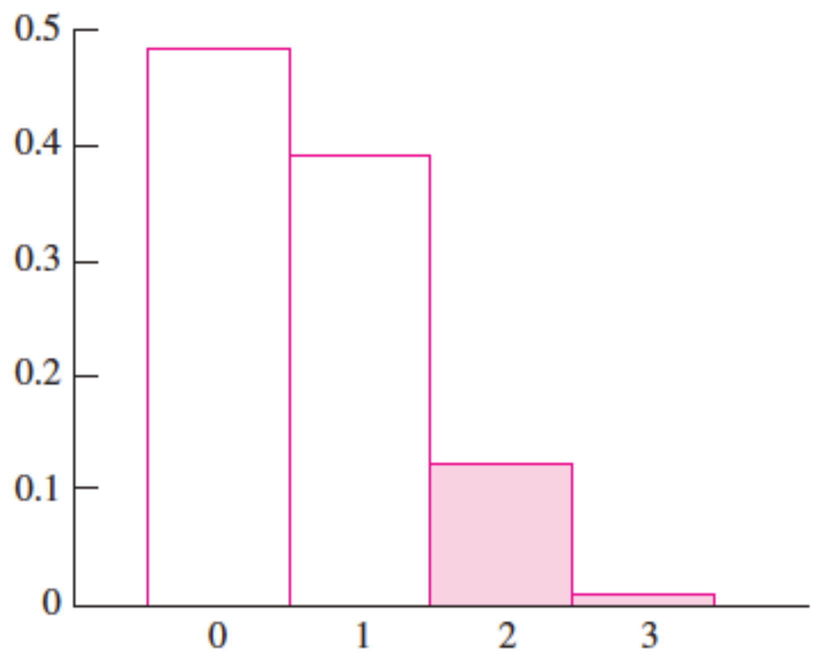
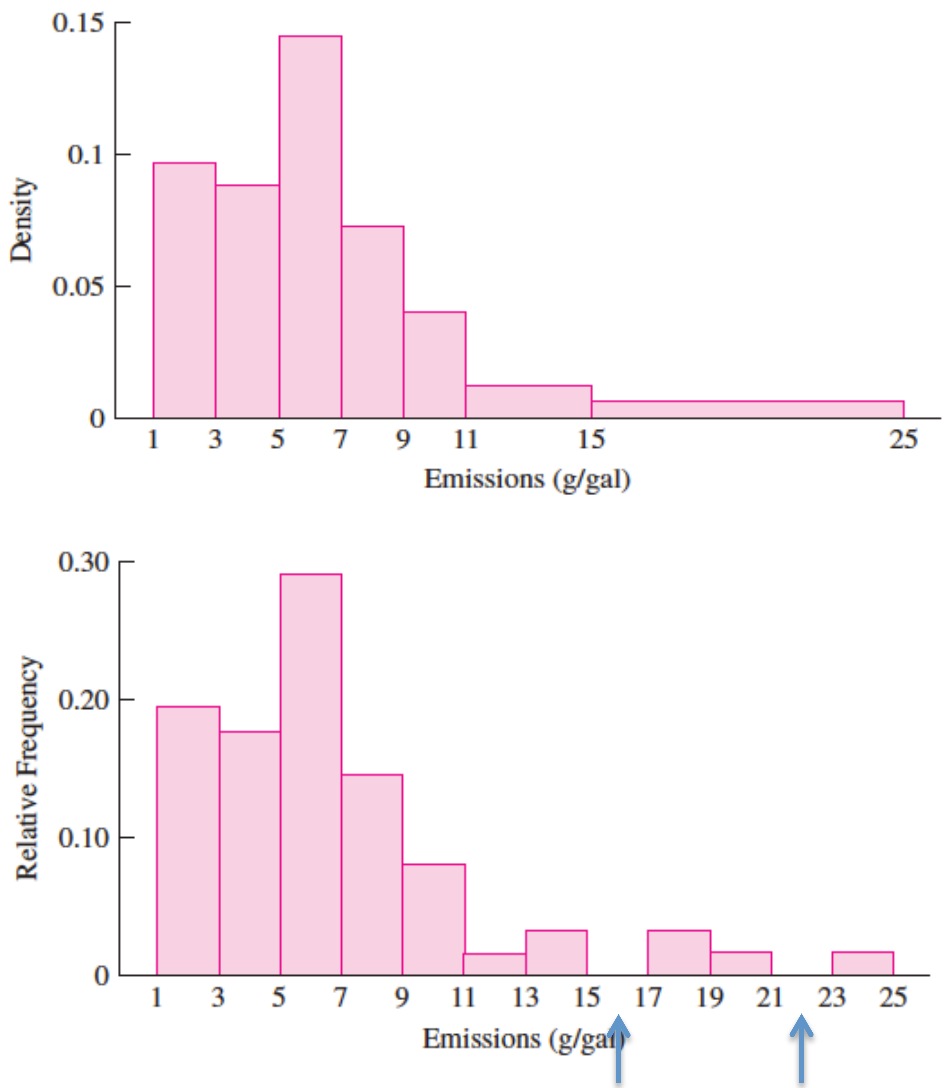


FIGURE 2.11 Probability histogram for X, the number of flaws in a randomly chosen piece of wire. The area corresponding to values of X greater than 1 is shaded. This area is equal to $P(X > 1)$.

Equal vs. Unequal Bin Widths



- Histograms of continuous measurement of vehicle emissions
- Using equal bin widths, there are intervals with no observations
- With larger sample, would likely observe values near 16 and 22

Histogram for Large Continuous Sample

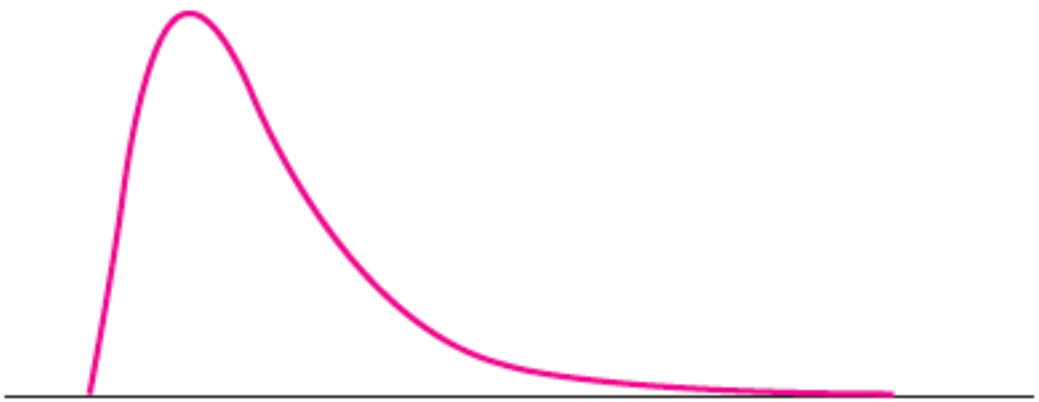
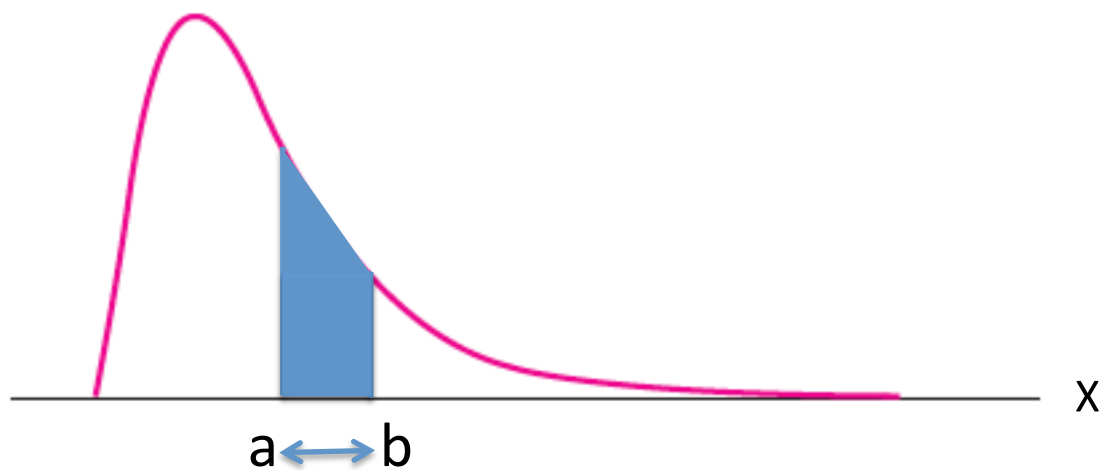


FIGURE 2.12 The histogram for a large continuous population could be drawn with extremely narrow rectangles and might look like this curve.

Continuous Random Variable

- The emission levels of a randomly chosen vehicle can be treated as a (continuous) random variable X
- In the probability histogram, the probability that X falls between any two values a and b is equal to the **area under the histogram** between a and b



Probability Density Function (Continuous)

- For continuous RVs, probabilities are given by areas under a curve
- The curve is called the **probability density function** or PDF (analogous to the pmf for discrete RVs)
- Integrate the curve between two points to find the proportion of values of the RV that lie in that interval
- Let X be a continuous RV, and a < b

$$P(a \le X \le b) = \int_a^b f(x)\,dx$$

Properties of the PDF

- The area does not depend on whether the endpoints of the interval are included

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

- The area under the entire curve must sum to one (the probability that X is between -∞ and +∞ is equal to one)

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

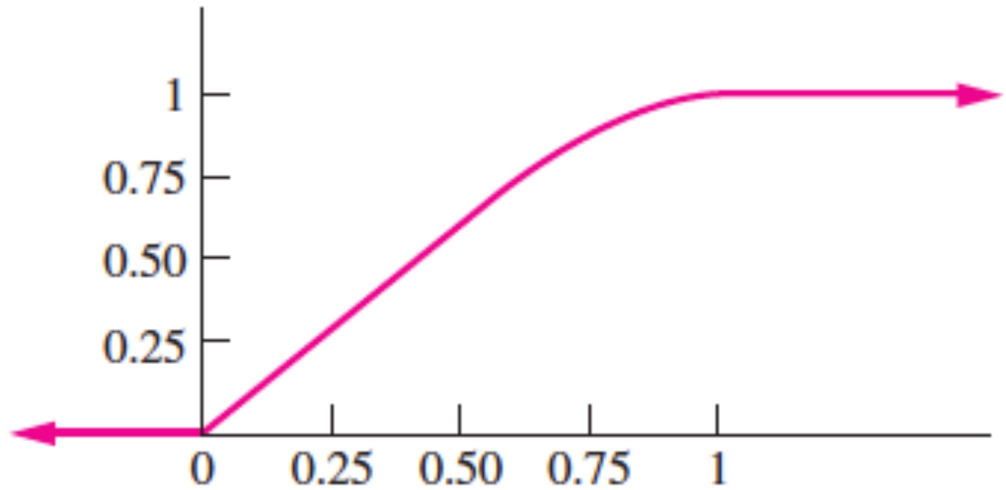
Cumulative Distribution Function (Continuous)

- For a continuous RV X that has probability density function f(x), the **cumulative distribution function** (CDF) is

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt = 1$$

- Analogous to the CDF of a discrete RV, except using an integral instead of summation

Continuity of the CDF



The CDF of a continuous RV is always continuous
(unlike the CDF of a discrete RV)

Example – Exercise 2.4.13

- Resistors labeled 100 ohms have true values between 80 and 120
- Let X be the resistance of a randomly chosen resistor
- The PDF of X is

$$f(x) = \begin{cases} \dfrac{x-80}{800} & 80 < x < 120 \\ 0 & \text{otherwise} \end{cases}$$

- What proportion of resistors have resistances more than 110?

Example – Exercise 2.4.13

- Resistors labeled 100 ohms have true values between 80 and 120
- Let X be the resistance of a randomly chosen resistor
- The PDF of X is

$$f(x) = \begin{cases} \dfrac{x-80}{800} & 80 < x < 120 \\ 0 & \text{otherwise} \end{cases}$$

- What proportion of resistors have resistances more than 110?

0.4375

Mean and Variance of Continuous RVs

- The **mean** (or expectation/expected value) of a continuous RV X is given by

$$\mu_X = \int_{-\infty}^{\infty} x f(x)\, dx$$

- The **variance** of a continuous RV X is given by

$$\sigma_X^2 = \int_{-\infty}^{\infty} (x - \mu_X)^2 f(x)\, dx$$

$$= \int_{-\infty}^{\infty} x^2 f(x)\, dx - \mu_X^2$$

Exercise 2.4.13 Continued

Find the mean and variance of X

$$f(x) = \begin{cases} \dfrac{x-80}{800} & 80 < x < 120 \\ 0 & \text{otherwise} \end{cases}$$

Exercise 2.4.13 Continued

Find the mean and variance of X

$$f(x) = \begin{cases} \dfrac{x-80}{800} & 80 < x < 120 \\ 0 & \text{otherwise} \end{cases}$$

E(X) = 106.67 ohms

Var(X) = 88.89 ohms^2

Percentiles

- The p^{th} percentile is the point x_p where p% of the values in the population are less than x_p ($0 \le p \le 100$)
- To find x_p solve the equation

$$F(x_p) = P(X \le x_p) = \int_{-\infty}^{x_p} f(x)\,dx = p/100$$

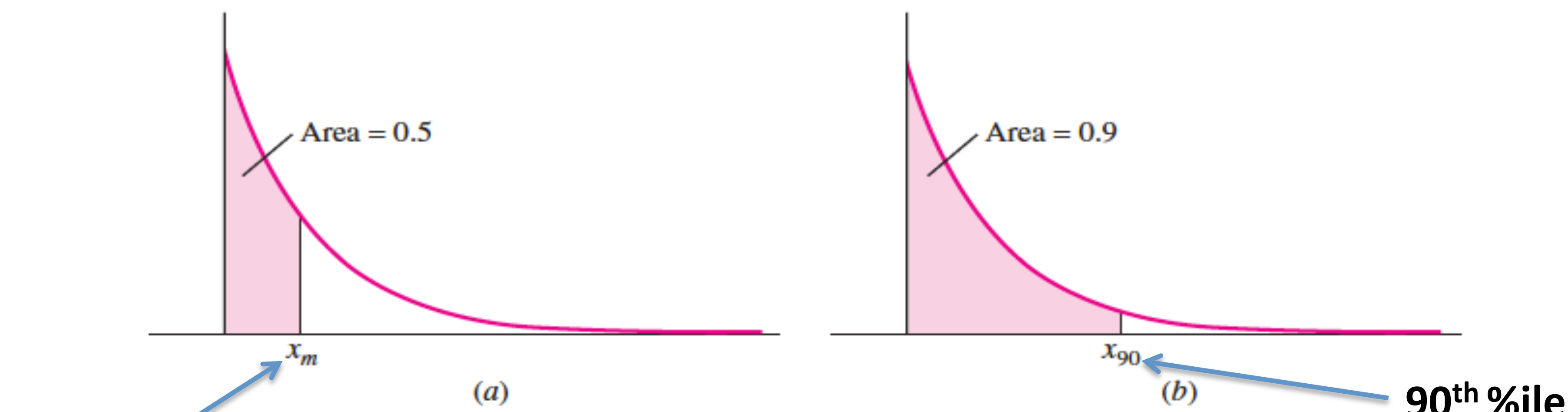


FIGURE 2.14 ***(a)*** Half of the population values are less than the median x_m. ***(b)*** Ninety percent of the population values are less than the 90th percentile x_{90}.

Example 2.45

A certain radioactive mass emits alpha particles from time to time. The time between emissions, in seconds, is random, with probability density function

$$f(x) = \begin{cases} 0.1e^{-0.1x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

Find the median time between emissions. Find the 60th percentile of the times.

Probability Bounds

- The standard deviation is a measure of the degree of spread around the center (mean)
- The probability that a random variable differs from its mean by *k* or more standard deviations is less than or equal to $1/k^2$
- This rule is called **Chebyshev's Inequality**

$$P(|X - \mu_X| \geq k\sigma) \leq \frac{1}{k^2}$$

Next

- Functions of random variables (2.5)