

Discrete and Continuous Random Variables



Summer 2003

Random Variables

- A *random variable* is a rule that assigns a numerical value to each possible outcome of a probabilistic experiment.
- We denote a random variable by a *capital letter* (such as "X")

Examples of random variables:

r.v. X: the age of a randomly selected student here today.

r.v. Y: the number of planes completed in the past week.

Discrete or Continuous

- A *discrete* r.v. can take only distinct, separate values
 - Examples?

- A *continuous* r.v. can take any value in some interval (low,high)
 - Examples?

Discrete Random Variables

■ A probability distribution for a *discrete* r.v. X consists of:

- Possible values x_1, x_2, \dots, x_n
- Corresponding probabilities p_1, p_2, \dots, p_n

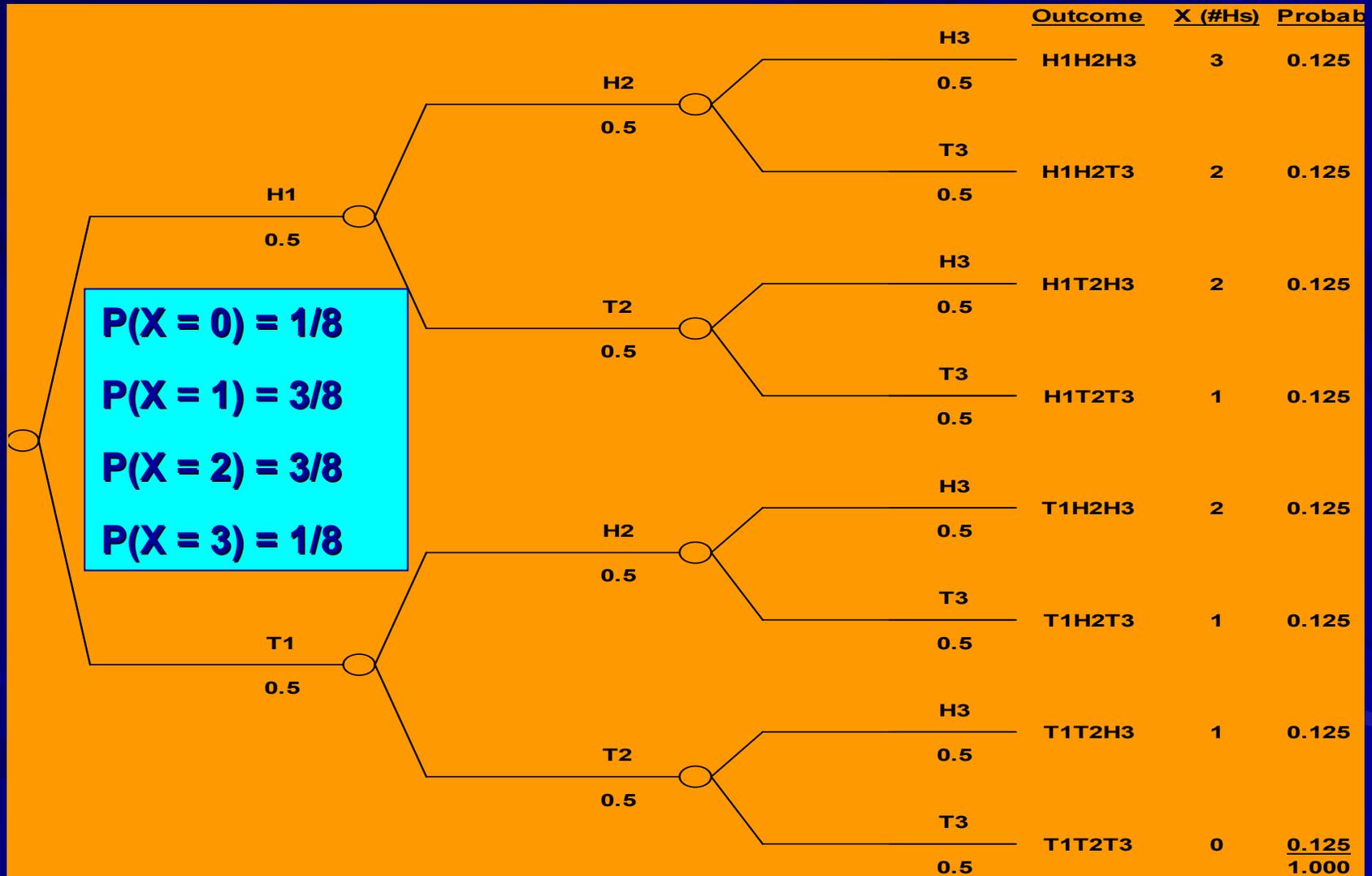
with the interpretation that

$$p(X = x_1) = p_1, \quad p(X = x_2) = p_2, \quad \dots, \quad p(X = x_n) = p_n$$

■ Note the following:

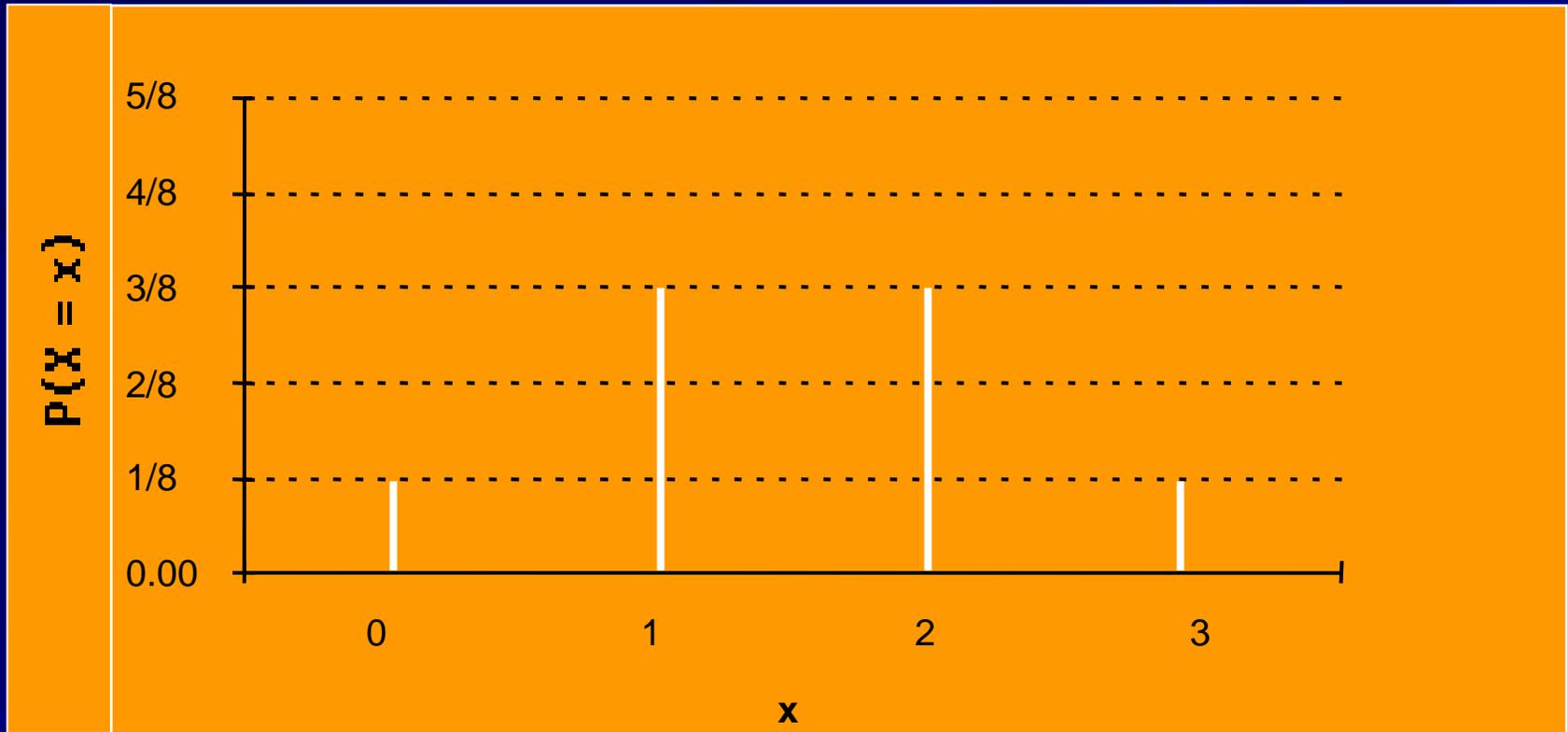
- Variable names are capital letters (e.g., X)
- Values of variables are lower case letters (e.g., x_1)
- Each $p_i \geq 0$ and $p_1 + p_2 + \dots + p_n = 1.0$

Probability tree and probability distribution for r.v. X (total # Heads in experiment 1)



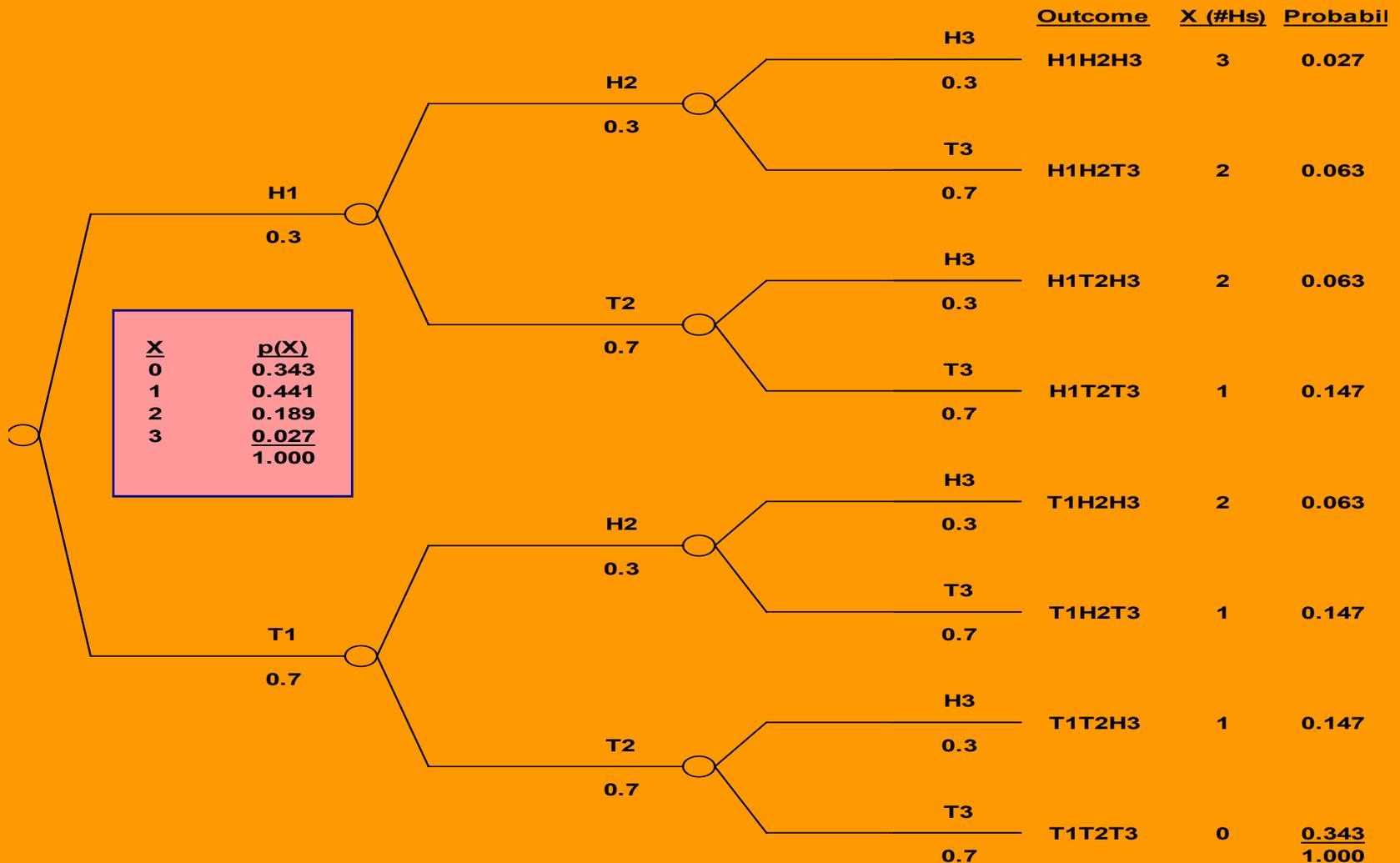
A histogram is a display of probabilities as a bar chart

r.v. $X =$ number of heads when tossing 3 unbiased coins



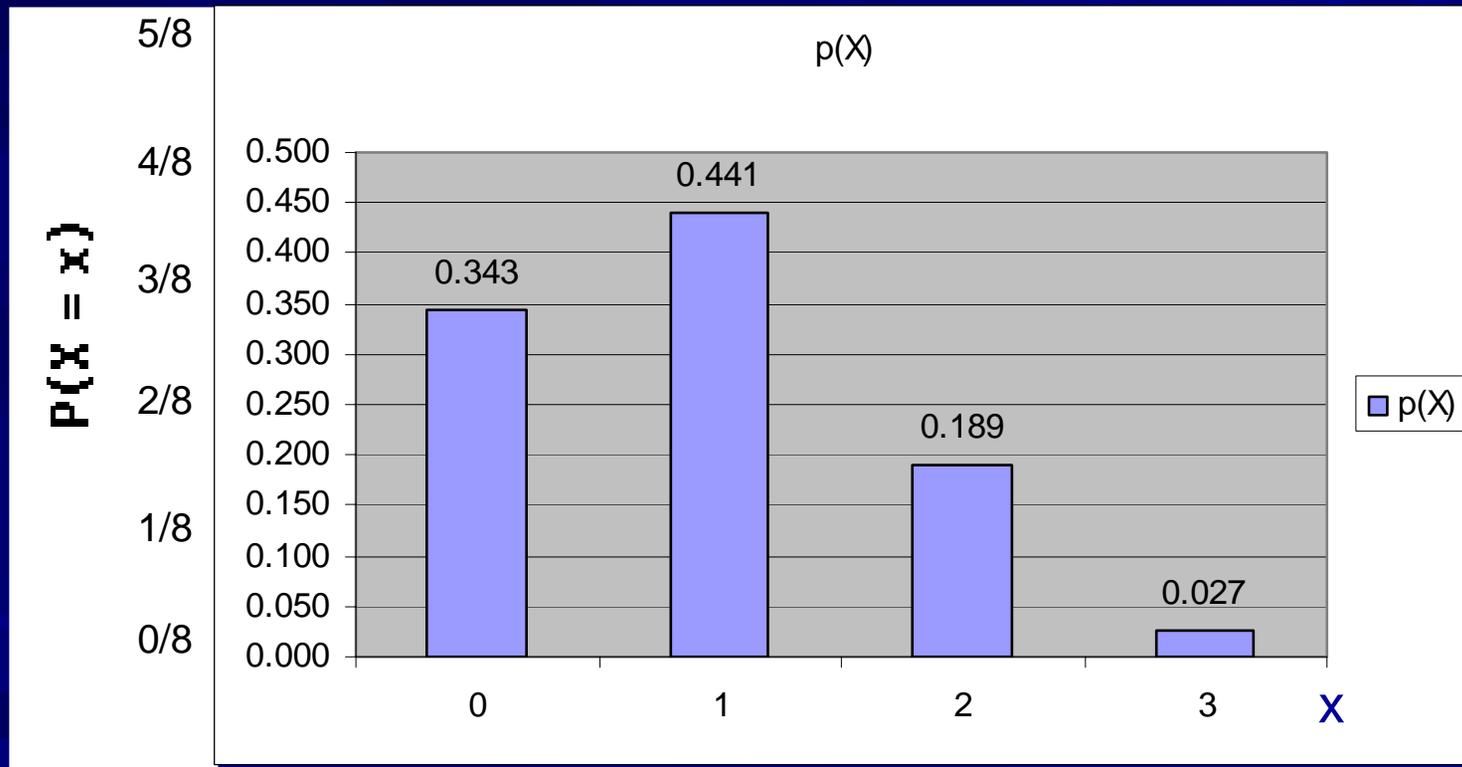
Now let's consider experiment 2: "number of heads when tossing 3 biased coins ($p(H) = 0.30$)". Again r.v. X : total number of heads obtained when performing experiment 2.....

Probability tree and probability distribution for r.v. X (total # Heads in experiment 2)



Histogram for experiment 2

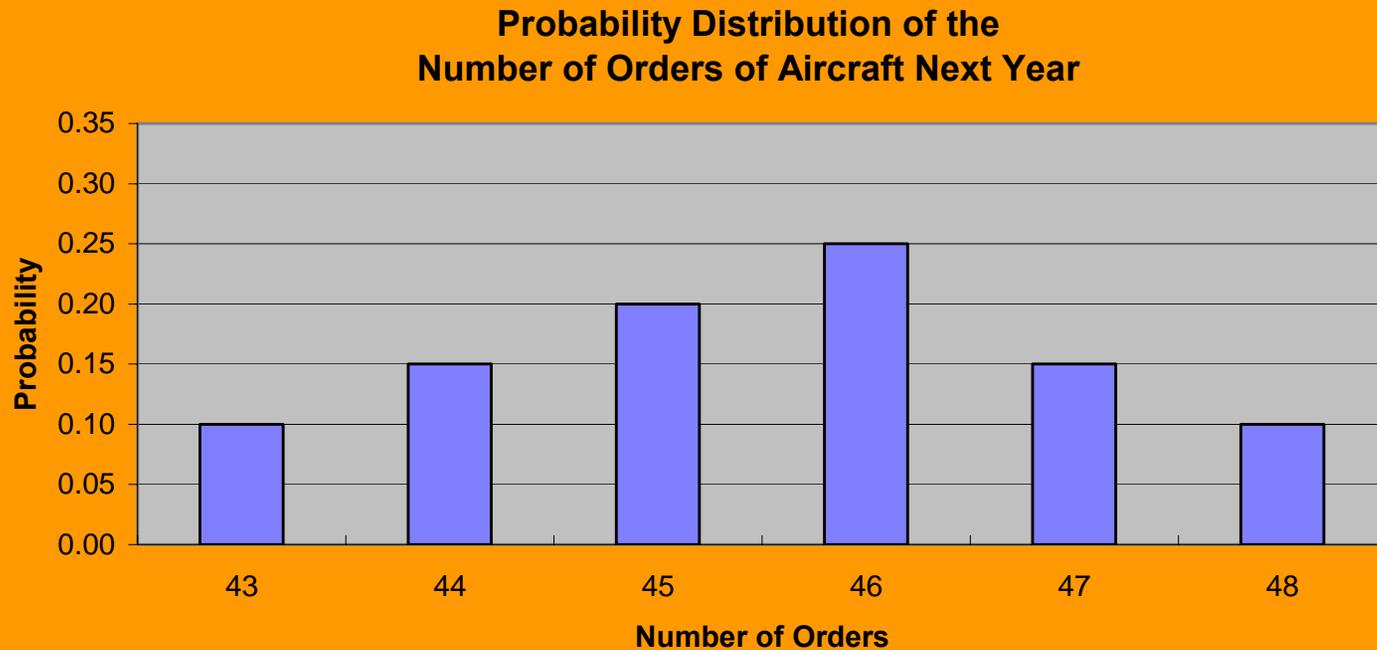
r.v. X = total number of heads when tossing 3 biased coins with $p(H) = 0.30$.



Example 2: Let X be the random variable that denotes the number of orders for aircraft for next year. Suppose that the number of orders for aircraft for next year is estimated to obey the following distribution:

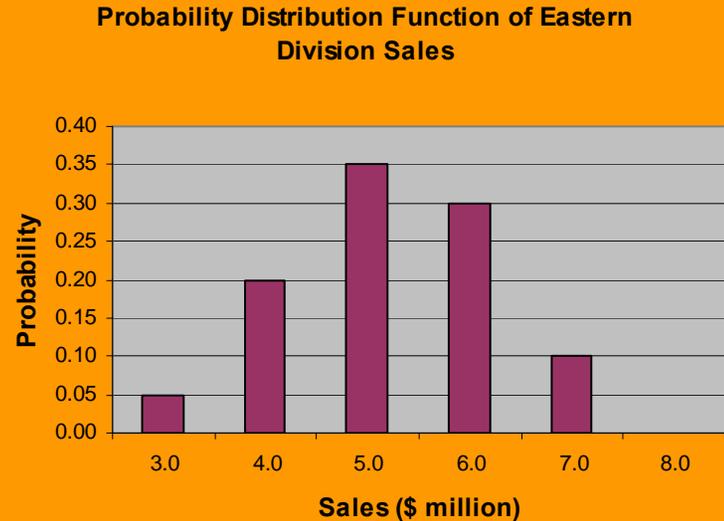
Orders for aircraft next year X_i	Probability p_i
42	0.05
43	0.10
44	0.15
45	0.20
46	0.25
47	0.15
48	0.10

A histogram is a display of probabilities as a bar chart



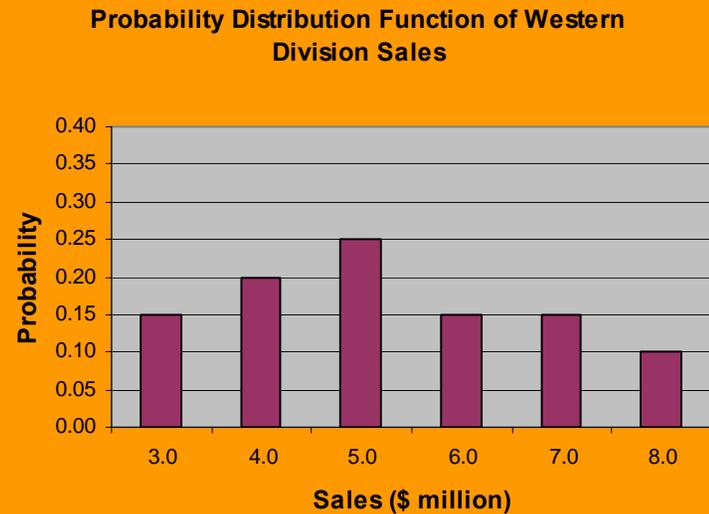
Eastern Division

Sales (\$ million)	Probability
3.0	0.05
4.0	0.20
5.0	0.35
6.0	0.30
7.0	0.10
8.0	0.00



Western Division

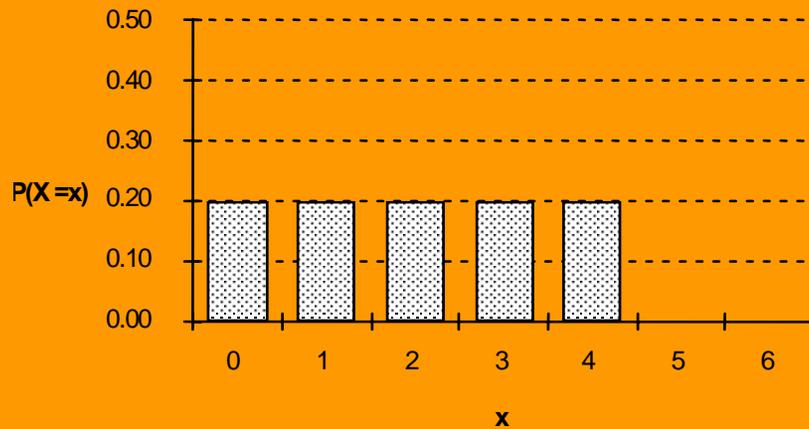
Sales (\$ million)	Probability
3.0	0.15
4.0	0.20
5.0	0.25
6.0	0.15
7.0	0.15
8.0	0.10



X and Y denote the sales next year in the eastern division and the western division of a company, respectively. X and Y obey the following distributions:

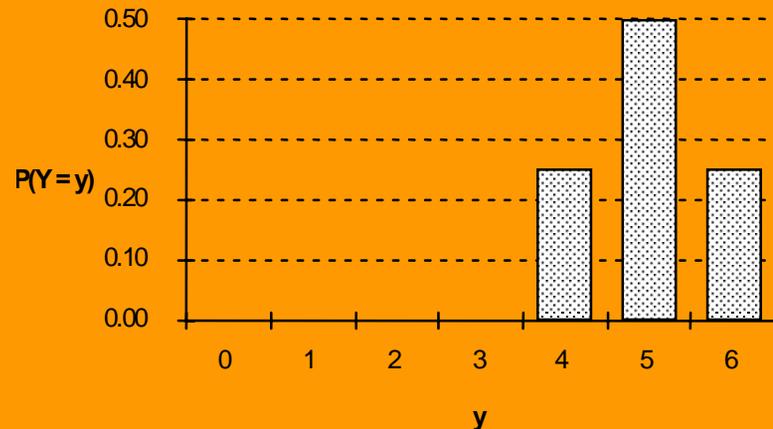
Consider two random variables X , Y , with the following histograms:

(A)



Random variable X

(B)



Random variable Y

How do we describe and compare X and Y ?

Summary Statistics for a r.v.: Three important measures

- Mean or Expected Value:

Represents "average" outcome; a measure of "central tendency"

$$E(X) = \mu_X = \sum_{i=1}^n P(X = x_i)x_i = \sum_{i=1}^n p_i x_i$$

- Variance:

Squared deviation around the mean; a measure of "spread"

$$\text{Var}(X) = \sigma^2_X = \sum_{i=1}^n P(X = x_i)(x_i - \mu_X)^2 = \sum_{i=1}^n p_i (x_i - \mu_X)^2$$

- Standard Deviation :

Square root of the variance. A measure of spread in the same units as the random variable X.

$$\text{SD}(X) = \sigma_X = \sqrt{\sigma_X^2}$$

Example: Let X be the number that comes up on a roll of one die.

→ Compute the mean, variance and standard deviation of X .

<u>Outcome</u>	<u>$P(X=x)$</u>
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

$$\mu_x = 1/6 * 1 + 1/6 * 2 + 1/6 * 3 + 1/6 * 4 + 1/6 * 5 + 1/6 * 6 = 3.5$$

$$\begin{aligned} (\sigma_x)^2 &= 1/6 * (1-3.5)^2 + 1/6 * (2-3.5)^2 + 1/6 * (3-3.5)^2 + 1/6 * (4-3.5)^2 \\ &+ 1/6 * (5-3.5)^2 + 1/6 * (6-3.5)^2 = 2.917 \end{aligned}$$

$$\sigma_x = \text{Sqrt}\{2.917\} = 1.708$$

Example 2:

→ Compute the mean, variance, standard deviation of X (eastern division sales) and Y (western division sales)

$$\mu_X = \$5.2 \text{ mil}$$

$$\mu_Y = \$5.25 \text{ mil}$$

$$\sigma_X^2 = 1.06 \text{ mil}^2$$

$$\sigma_Y^2 = 2.3875 \text{ mil}^2$$

$$\sigma_X = \$1.029 \text{ mil}$$

$$\sigma_Y = \$1.545 \text{ mil}$$

Continuous Random Variables

A *continuous* random variable can take any value in some interval

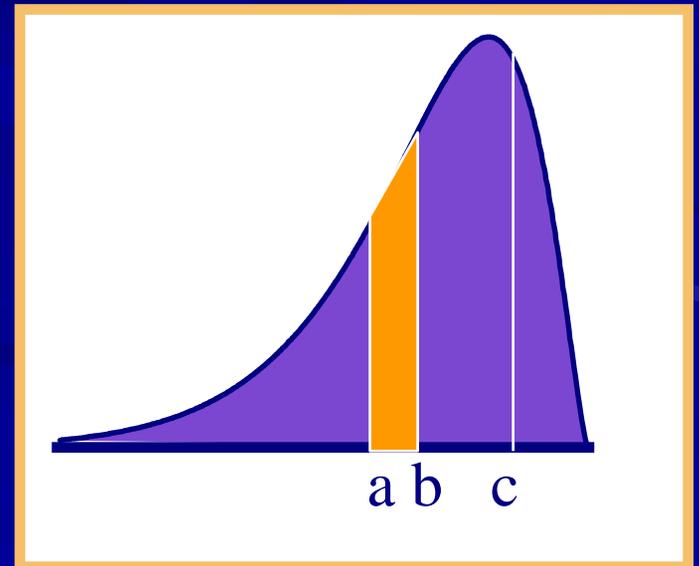
Example: X = time a customer spends waiting in line at the store

- “Infinite” number of possible values for the random variable.
How can we describe a probability distribution?
(We can no longer list the p_i 's and x_i 's!)
- For a continuous random variable, questions are phrased in terms of a range of values.

Example: We might talk about the event that a customer waits between 5.0 and 10.0 minutes, and not about the event that a customer waits exactly 5.25 minutes! (why?)

Continuous Probability Distributions

- Continuous R.V.'s have continuous probability distributions known also as the probability density function (PDF)
- Since a continuous R.V. X can take an **infinite** number of values on an interval, the probability that a continuous R.V. X takes any single given value is **zero**: $P(X=c)=0$
- Probabilities for a continuous RV X are calculated for a range of values: $P(a \leq X \leq b)$
- $P(a \leq X \leq b)$ is the area under the probability distribution function, $f(x)$, for continuous R.V. X .
- The total area under $f(x)$ is 1.0



The Uniform Distribution

X is *uniform* on $[a,b]$ if X is equally likely to take any value in the range from a to b .

Example: Suppose that transit time of the subway between Alewife Station and Downtown Crossing is uniformly distributed between 10.0 and 20.0 minutes.

a) What is the mean transit time?

$$E(X) = \mu_x = ?$$

b) What is the probability that the transit time exceeds 12.0 minutes?

$$P(X \geq 12.0) = ?$$

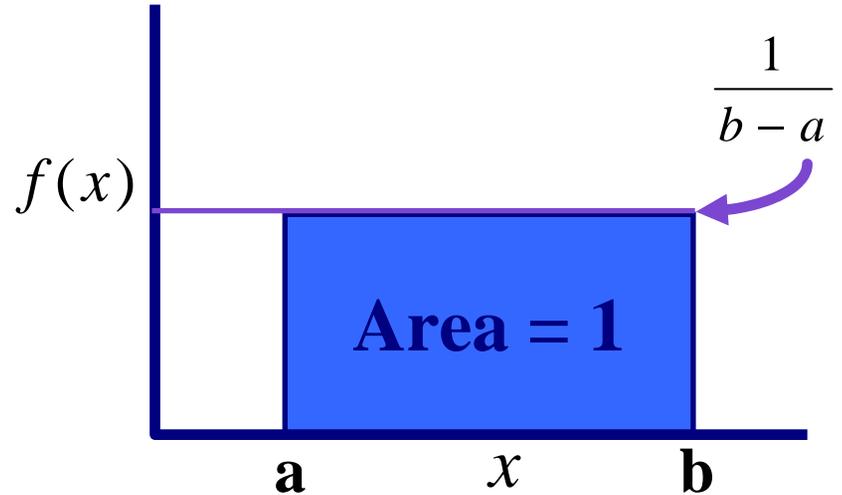
c) What is the probability that the transit time is between 14.0 and 18.0 minutes?

$$P(14 \leq X \leq 18) = ?$$

Uniform Distribution

- A continuous R.V. X is uniform in the interval $[a-b]$ if it is equally likely to take any value in the interval $[a,b]$. Thus $f(x)$ for a continuous uniform R.V. is a horizontal line (i.e., a constant). $f(x)=1/(b-a)$ so that the area under the curve is 1.0.

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq x \leq b \\ 0 & \text{for all other values} \end{cases}$$



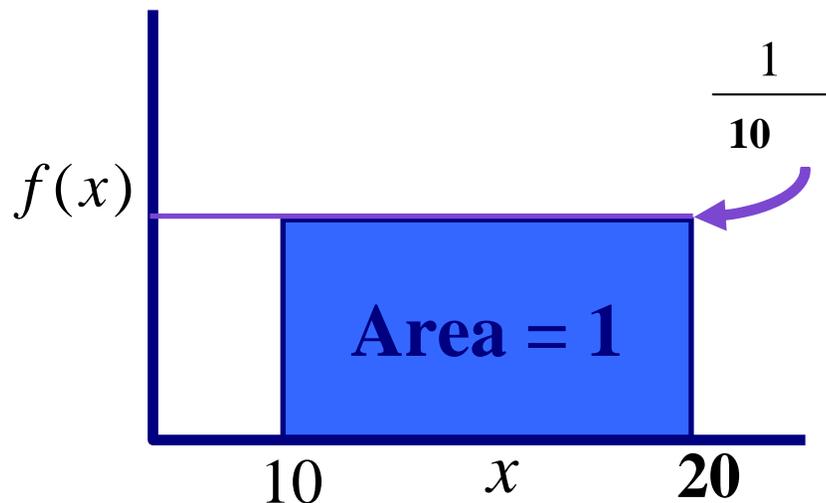
Uniform Distribution

- transit time of the subway between Alewife Station and Downtown Crossing is uniformly distributed between 10.0 and 20.0 minutes.

$$f(x) = \begin{cases} \frac{1}{20-10} & \text{for } 10 \leq x \leq 20 \\ 0 & \text{for all other values} \end{cases}$$

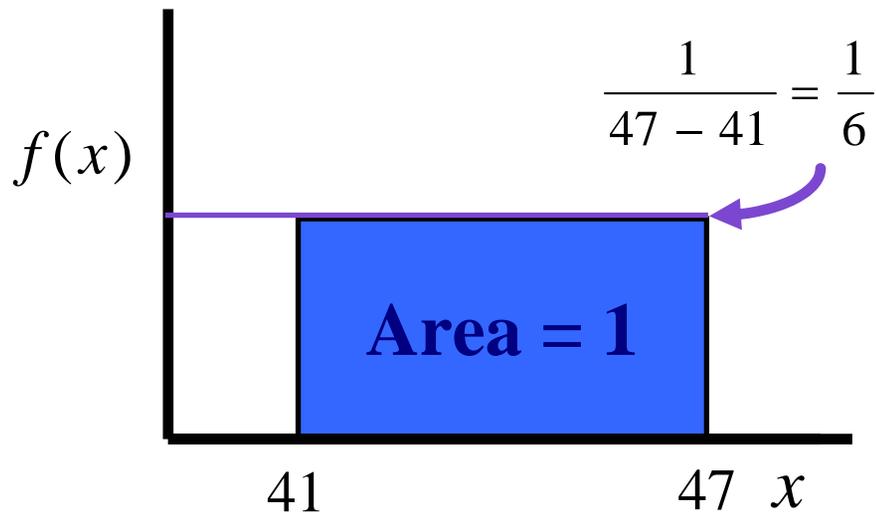
$$P(X \geq 12.0) = 0.80$$

$$P(14 \leq X \leq 18) = 0.40$$



Lot Weights in a Warehouse are Uniformly distributed Between 41 and 47 lbs.

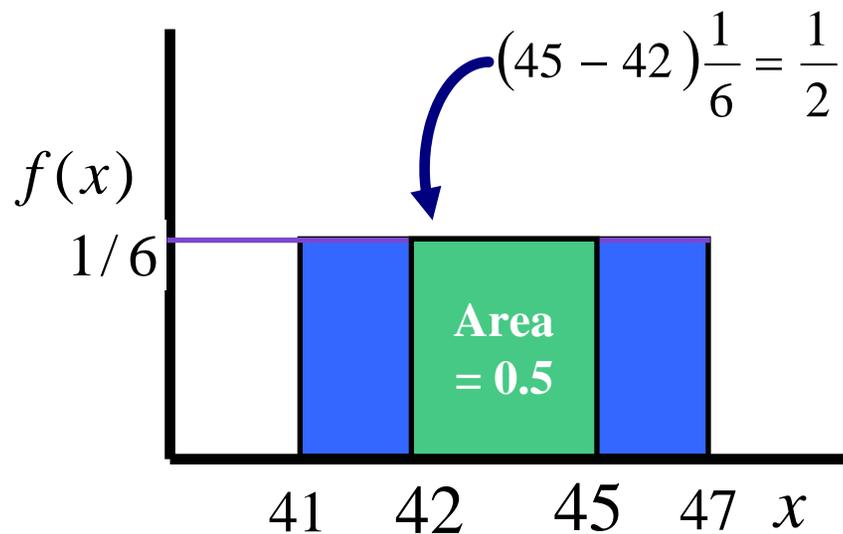
$$f(x) = \begin{cases} \frac{1}{47-41} & \text{for } 41 \leq x \leq 47 \\ 0 & \text{for all other values} \end{cases}$$



What is the probability that the weight of a randomly selected lot is between 42 and 45 lbs?

$$P(x_1 \leq X \leq x_2) = \frac{1}{b-a}(x_2 - x_1)$$

$$P(42 \leq X \leq 45) = (45 - 42) \frac{1}{6} = \frac{1}{2}$$



Uniform Distribution

Mean and Standard Deviation

Mean

$$\mu = \frac{a + b}{2}$$

Mean

$$\mu = \frac{41 + 47}{2} = \frac{88}{2} = 44$$

Standard Deviation

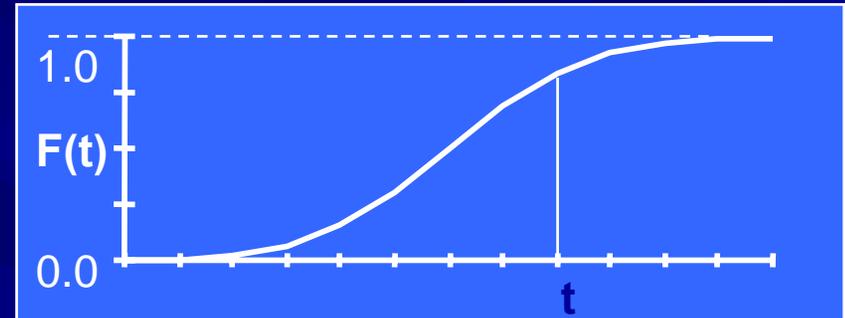
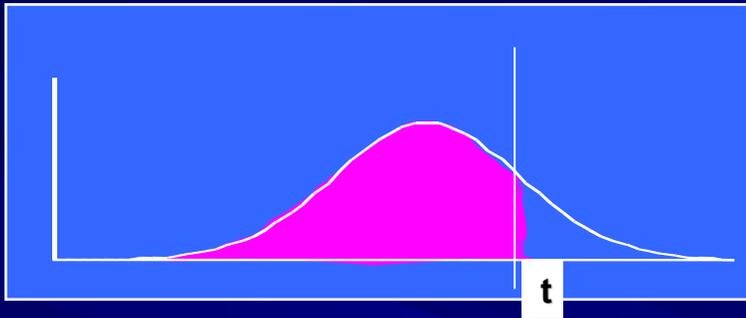
$$\sigma = \frac{b - a}{\sqrt{12}}$$

Standard Deviation

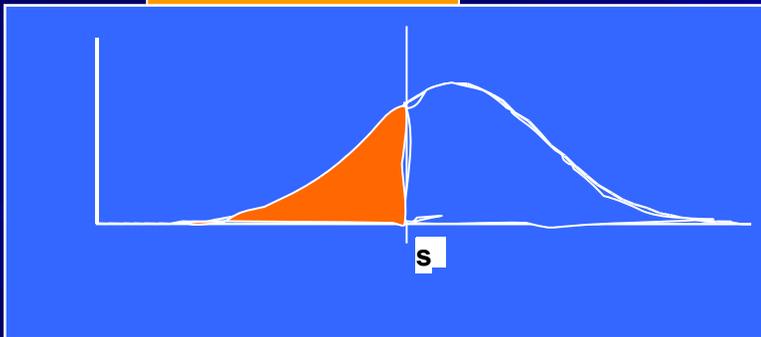
$$\sigma = \frac{47 - 41}{\sqrt{12}} = \frac{6}{3.464} = 1.732$$

Cumulative Distribution Function (CDF)

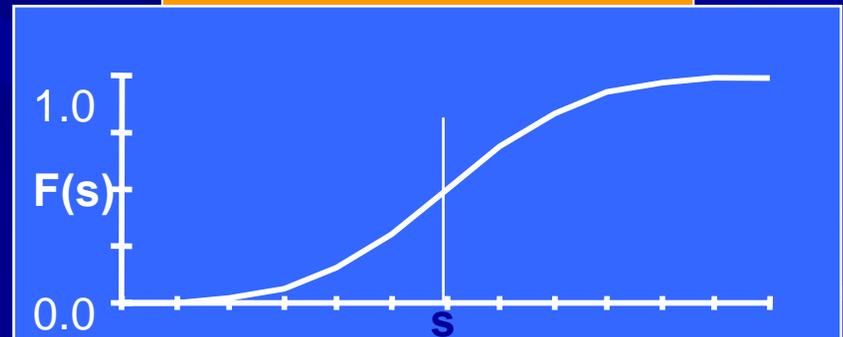
$$P(X \leq t) = \int_{-\infty}^t f(x) dx = F(t)$$



$$P(X \leq s)$$



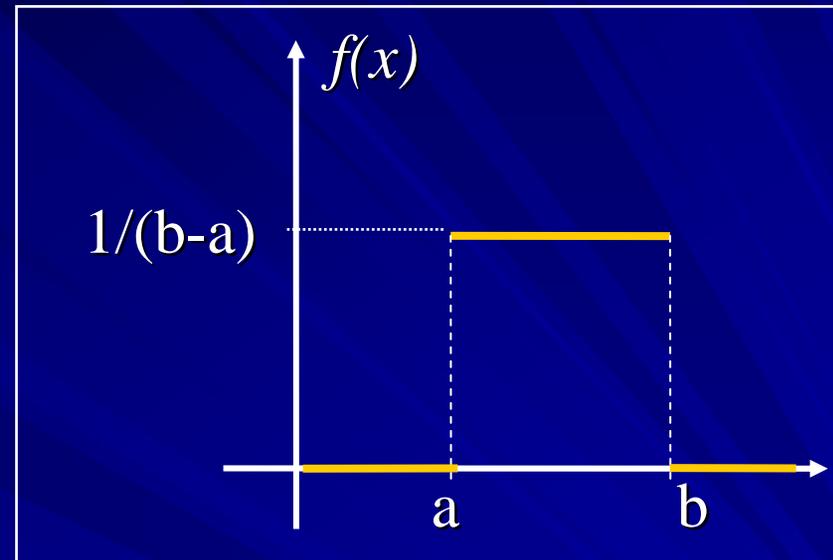
$$F(s) = P(X \leq s)$$



$$P(s \leq X \leq t) = P(X \leq t) - P(X < s) = F(t) - F(s)$$

Let X be uniformly distributed on $[a,b]$. Then density function of X is:

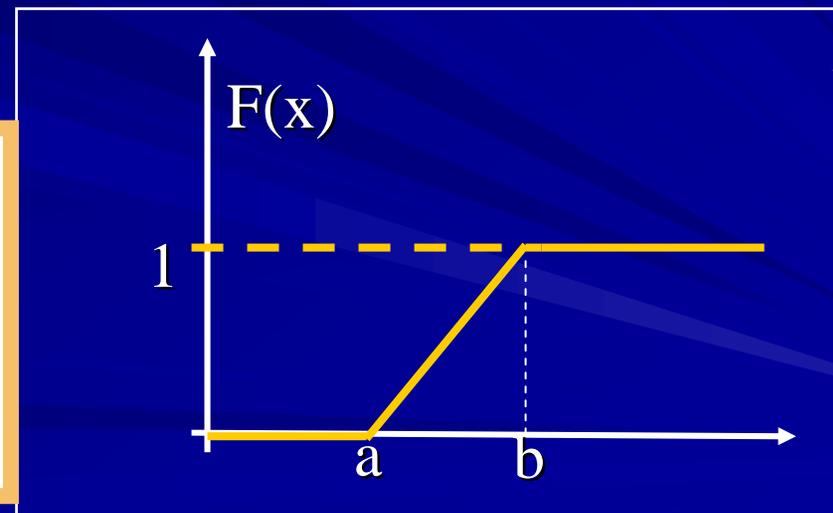
$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq x \leq b \\ 0 & \text{for all other values} \end{cases}$$



Find the CDF:

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

$$F(x) = \begin{cases} \frac{1}{(b-a)}x - \frac{1}{(b-a)}a & \text{for } a \leq x \leq b \\ 0 & \text{for } x < a \\ 1 & \text{for } x > b \end{cases}$$



Summary

- Discrete and continuous random variables have to be understood differently
- Histograms are very useful for discrete variables
- Next session we will talk about “binomial” distributions
- Uniform continuous random variables are a good place to start for our later work with normal distributions etc.