MATH1. Part II

Probability and Statistics

Chapter 3

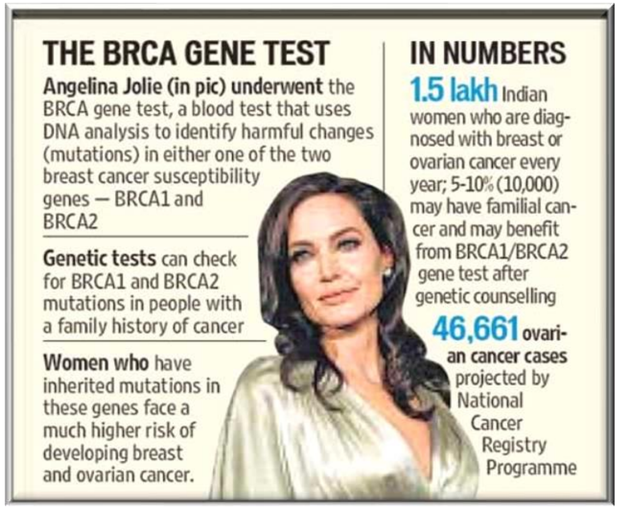
Probability and the Binomial Distribution

3.1 Probability and the Life Sciences

Probability, or chance, plays an important role in scientific thinking about living systems.

- Probability models allow us to quantify how likely, or unlikely, an experimental result is, given certain modeling assumptions.
- Some biological processes are affected directly by chance: occurrence of mutations
- The results of an experiment are always somewhat affected by chance: chance fluctuations in environmental conditions

1. Weather forecasting



THE BRCA GENE TEST

Angelina Jolie (in pic) underwent the BRCA gene test, a blood test that uses DNA analysis to identify harmful changes (mutations) in either one of the two breast cancer susceptibility genes — BRCA1 and BRCA2

Genetic tests can check for BRCA1 and BRCA2 mutations in people with a family history of cancer

Women who have inherited mutations in these genes face a much higher risk of developing breast and ovarian cancer.

IN NUMBERS

1.5 lakh Indian women who are diagnosed with breast or ovarian cancer every year; 5-10% (10,000) may have familial cancer and may benefit from BRCA1/BRCA2 gene test after genetic counselling

46,661 ovarian cancer cases projected by National Cancer Registry Programme

2. Genetic Testing/Genetic Counseling

3.2 Introduction to Probability

Probability

- We can speak meaningfully about a probability only in the context of a chance operation—that is, an operation whose outcome is NOT deterministic.
- **Chance operation**: defined in such a way that each time the chance operation is performed, the event E either occurs or does not occur.

- **Probability:** a numerical quantity that expresses the likelihood of an event E.
- The probability of an event E is written as **Pr{*E*}**.
- The probability Pr{E} is always a number between 0 and 1, inclusive.

  $\mathbf{0 \le Pr\{\mathit{E}\} \le 1}$, 0 = certain non-occurrence, 1 = certain occurrence

3.2 Introduction to Probability

Probability

Example 3.2.1 Coin Tossing

- What is the chance operation?
- Define the event *E*
- If the coin is <u>equally</u> likely to fall heads or tails, then what is the value of Pr{*E*}?

3.2 Introduction to Probability

Probability

Example 3.2.1 Coin Tossing

- What is the chance operation?
  - Tossing a coin
- Define the event *E*
  - *E:* Heads or *E*: Tails

- If the coin is equally likely to fall heads or tails, then what is the value of Pr{*E*}?
  - Pr{*E*} = 1/2 = 0.5
    - Such an ideal coin is called a "fair" coin.
    - If the coin is not fair (perhaps because it is slightly bent), then Pr{E} will be some value other than 0.5, for instance, Pr{E} = 0.6

3.2 Introduction to Probability

Probability

Example 3.2.3 Sampling Fruitflies

- In a fruitfly population, 30% of the individuals are black because of a mutation, while 70% of the individuals have the normal gray body color. Suppose one fly is chosen at random from the population.
- What is the chance operation?
- Define the event E
- What is the probability of getting a black fly?

3.2 Introduction to Probability

Probability

Example 3.2.3 Sampling Fruitflies

- In a fruitfly population, 30% of the individuals are black because of a mutation, while 70% of the individuals have the normal gray body color. Suppose one fly is chosen at random from the population.
- What is the chance operation?
  - Chose one fly is chosen at random from the population
- Define the event E
  - Event E: Sampled fly is black
- What is the probability of getting a black fly?
  - Then the probability that a black fly is chosen Pr{E} = 0.3

3.2 Introduction to Probability

Probability

- The preceding example illustrates the basic relationship between probability and random sampling:

*The probability that a randomly chosen individual has a certain characteristic is equal to the **proportion** of population members with the characteristic.*

Frequency interpretation of probability

- The probability Pr{E} is interpreted as: the relative frequency of occurrence of *E* in an indefinitely long series of repetitions of the chance operation.

$$\Pr\{E\} \leftrightarrow \frac{\#\text{ of times } E \text{ occurs}}{\#\text{ of times chance operation is repeated}}$$

3.2 Introduction to Probability

Frequency interpretation of probability

- The probability Pr{E} is interpreted as: the relative frequency of occurrence of *E* in an indefinitely long series of repetitions of the chance operation.

$$\Pr\{E\} \leftrightarrow \frac{\text{\# of times } E \text{ occurs}}{\text{\# of times chance operation is repeated}}$$

Example 3.2.1 Coin Tossing

- Consider again the chance operation of tossing a coin, and the event *E*: Heads
- If the coin is fair, then Pr {E} = 0.5
- Explain the meaning of Pr{E} by the frequency interpretation of probability.

3.2 Introduction to Probability

Frequency interpretation of probability

- The probability Pr{E} is interpreted as: the relative frequency of occurrence of *E* in an indefinitely long series of repetitions of the chance operation.

$$\Pr\{E\} \leftrightarrow \frac{\#\text{ of times } E \text{ occurs}}{\#\text{ of times chance operation is repeated}}$$

Example 3.2.1 Coin Tossing

- Consider again the chance operation of tossing a coin, and the event *E*: Heads
- If the coin is fair, then Pr {E} = 0.5
- Explain the meaning of Pr{E} by the frequency interpretation of probability.

$$\text{Pr }\{\text{E}\} = 0.5 \quad \longleftrightarrow \quad \frac{\#\ of\ Heads}{\#\ of\ Tosses}$$

  - The arrow in the preceding expression indicates that, in an infinitely long series of tosses of a fair coin, we expect to get heads about 50% of the time.

3.2 Introduction to Probability

Independent events

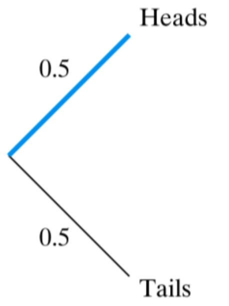
- Two events are said to be **independent** if knowing that one of them occurred does **not** change the probability of the other one occurring.

Probability trees

- A probability tree provides a convenient way to break a problem into parts and to organize the information available.

Example 3.2.7 Coin Tossing (2 chance operation)

- If a fair coin is tossed twice, what is the probability of getting 2 heads?
  - Event E: heads on both tosses
  - Fair coin: the probability of heads is 0.5 on each toss.
  - The first part of a probability tree for this scenario shows that there are two possible outcomes for the first toss and that they have probability 0.5 each.



3.2 Introduction to Probability

Probability trees

Example 3.2.7 Coin Tossing (2 chance operation - continued)

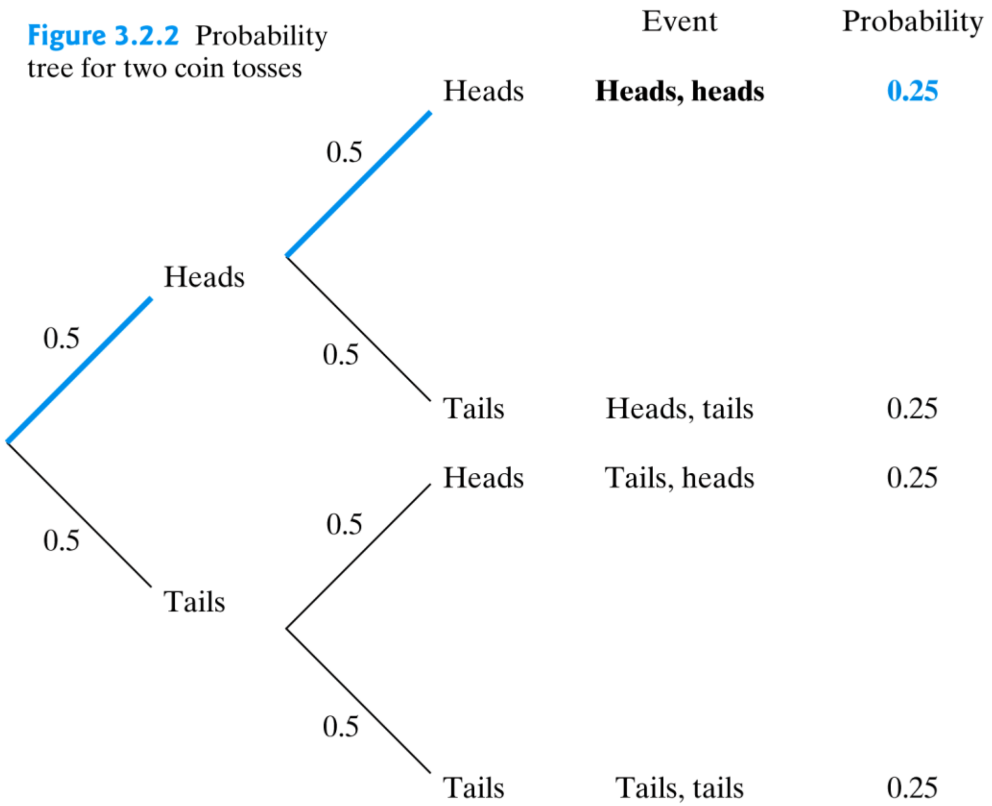


Figure 3.2.2 Probability tree for two coin tosses

- The second part of the probability tree had the same structure as the first part for <u>independent events</u>.
- For either outcome of the 1st toss, the 2nd toss can be either heads or tails, again with probabilities 0.5 each.
- To find the probability of getting heads on both tosses, we consider the path through the tree that produces this event.
- We multiply together the probabilities that we encounter along the path.
- Pr {heads on both tosses} = 0.5 x 0.5 = 0.25

3.2 Introduction to Probability

Probability trees - Combination of probabilities

Example 3.2.8 Sampling Fruitflies

- In a fruitfly population, 30% of the individuals are black because of a mutation, while 70% of the individuals have the normal gray body color. Suppose that two flies are randomly chosen from the population.
- What is the probability that both flies are the same color?

Hint:

– Event E: Both flies in the sample are the same color

– Pr {Both flies in the sample are the same color}

= Pr{getting two black flies} + Pr{getting two gray flies }

3.2 Introduction to Probability

Probability trees - Combination of probabilities

Example 3.2.8 Sampling Fruitflies (continued)

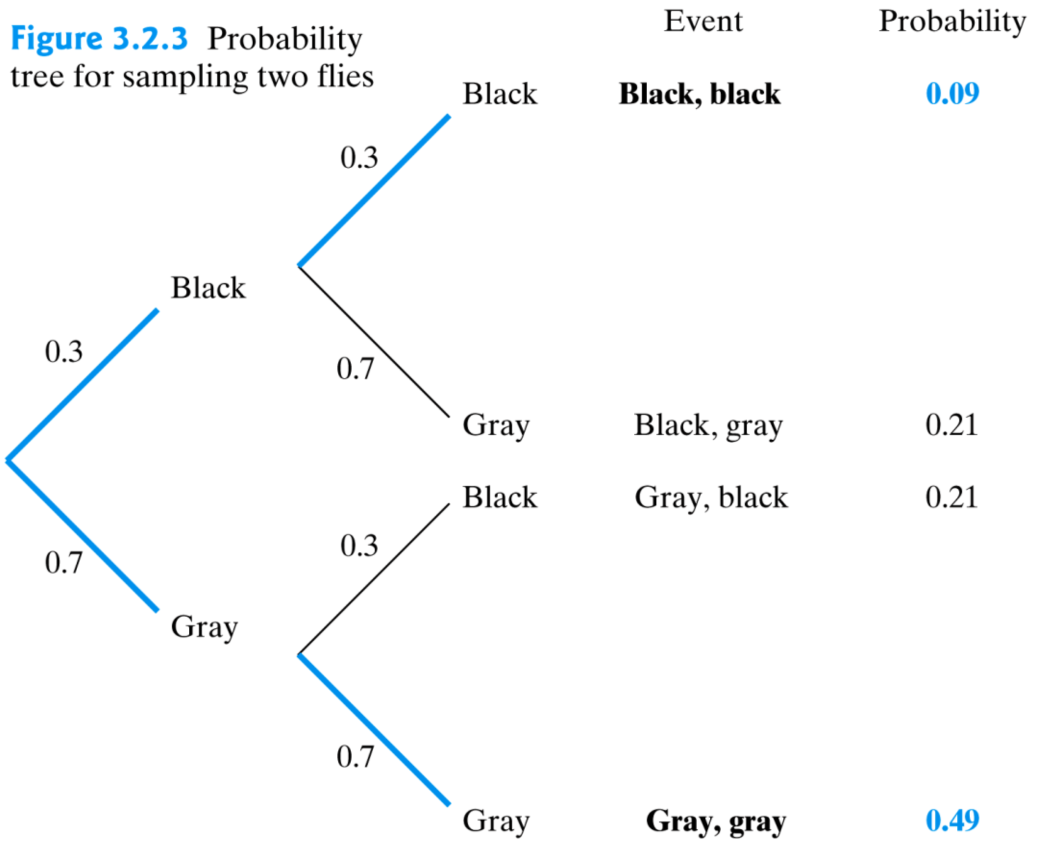


Figure 3.2.3 Probability tree for sampling two flies

- What is the probability that both flies are the same color?

Pr {Both flies in the sample are the same color}

= Pr{getting two black flies} + Pr{getting two gray flies }

= 0.3 x 0.3 + 0.7 x 0.7

= 0.09 + 0.49

= 0.58

3.2 Introduction to Probability

Probability trees - Combination of probabilities

Example 3.2.9 Nitric Oxide

- One treatment for hypoxic respiratory failure is to have the newborn inhale nitric oxide.
- To test the effectiveness of nitric oxide treatment, newborns suffering hypoxic respiratory failure were assigned at random to either be given nitric oxide or a control group.
- In the treatment group 45.6% of the newborns had a negative outcome, meaning that either they needed ECMO or that they died. In the control group, 63.6% of the newborns had a negative outcome.
- What is the probability of having a negative outcome?

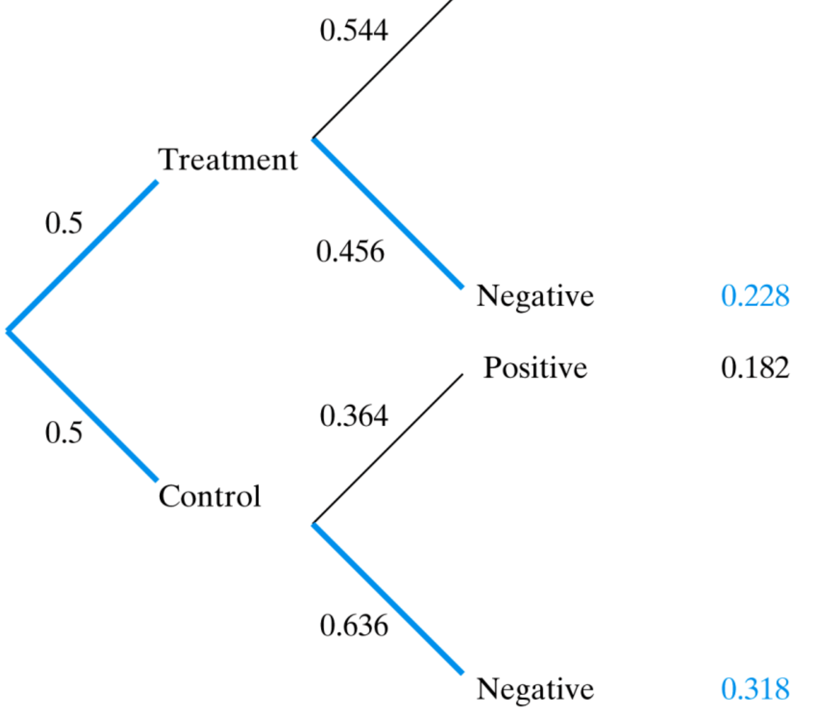

10 min Break

3.2 Introduction to Probability

Probability trees - Combination of probabilities

Example 3.2.9 Nitric Oxide (continued)

Figure 3.2.4 Probability tree for nitric oxide example



- What is the probability of having a negative outcome?

— Event E: negative outcome

— Pr {negative outcome}

= Pr{negative outcome in control group}

+ Pr{negative outcome w/Nitric Oxide}

= 0.5 x 0.456 + 0.5 x 0.636

= 0.228 + 0.318

= 0.546

3.3 Probability Rules

Basic Rules

- **Rule (1):** The probability of an event E is always between 0 and 1. That is,

  $0 \le \Pr\{E\} \le 1$, 0 = certain non-occurrence, 1 = certain occurrence

- **Rule (2):** The sum of the probabilities of all possible events equals 1.That is, if the set of possible events is E_1 ,E_2 , . . . , E_k , then

  $$\sum_{i=1}^{k} \Pr\{Ei\} = 1$$

- **Rule (3):** The probability that an <u>event E does not happen</u>, denoted by <u>E^C</u> ,is one minus the probability that the event happens. That is,

  $\Pr\{E^C\} = 1 - \Pr\{E\}$, We refer to E^C as the complement of E.

3.3 Probability Rules

Basic Rules

Example 3.3.1 Blood Type

- In the United States, 44% of the population has type O blood, 42% has type A, 10% has type B, and 4% has type AB.
- Consider choosing someone at random and determining the person's blood type.
  - **Rule (1): $0 \leq Pr\{E\} \leq 1$, 0 = certain non-occurrence, 1 = certain occurrence**
    - Pr{O} = 0.44, Pr{A} = 0.42, Pr{B} = 0.10, Pr{AB} = 0.04.

      ** The probability of a given blood type will correspond to the population percentage.*
  - **Rule (2): $\sum_{i=1}^{k} \mathbf{Pr}(\boldsymbol{Ei}) = \mathbf{1}$**
    - Pr{O} + Pr{A} + Pr{B} + Pr{AB} = 0.44 + 0.42 + 0.10 + 0.04 = 1
  - **Rule (3): $Pr\{E^C\} = 1 - Pr\{E\}$**
    - $Pr\{O^C\} = 1 - Pr\{O\} = Pr\{A\} + Pr\{B\} + Pr\{AB\} = 0.56$

3.3 Probability Rules

Venn Diagram

- We often want to discuss two or more events at once; to do this we will find some terminology to be helpful.
- We say that two events are **disjoint**, if they cannot occur simultaneously.
- We say that two events are NOT **disjoint**, if they can occur simultaneously.

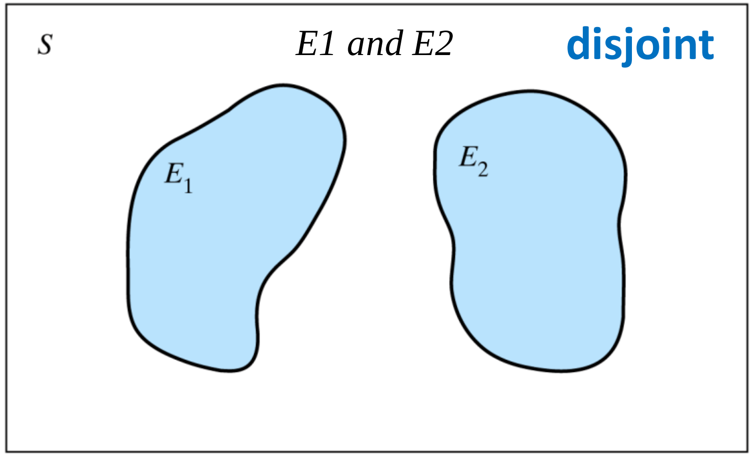


Figure 3.3.1 Venn diagram showing two disjoint events

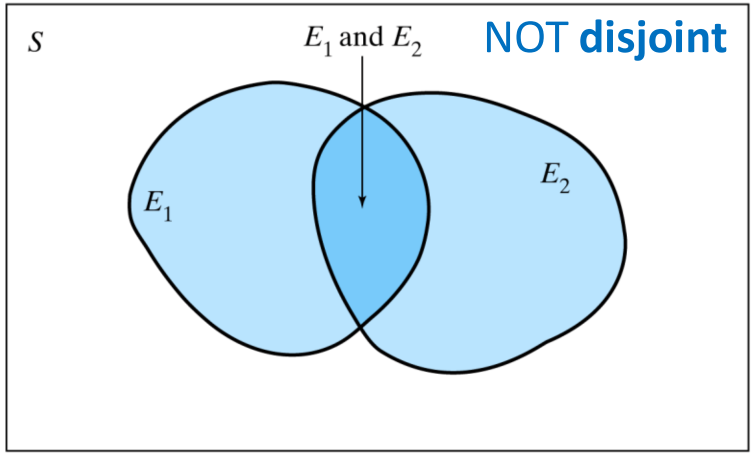
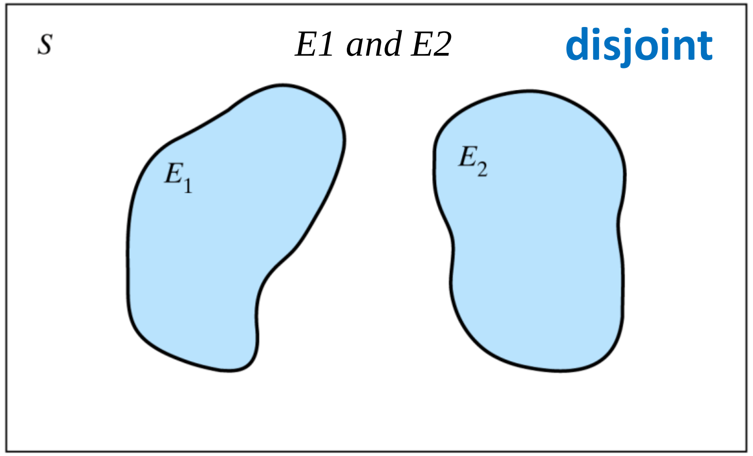


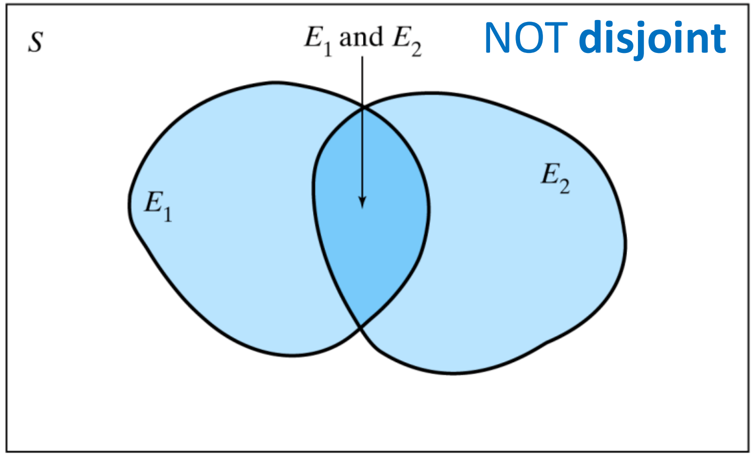
Figure 3.3.2 Venn diagram showing union (total shaded area) and intersection (middle area) of two events

3.3 Probability Rules

Venn Diagram

- The **union** of two events is the event that one or the other occurs or both occur.
- The **intersection** of two events is the event that they both occur.





- **Figure 3.3.1** is a Venn diagram that depicts a sample space S of all possible outcomes as a rectangle with **two disjoint events** depicted as nonoverlapping regions.
- If two events are disjoint, then the probability of their **union** is the sum of their individual probabilities.

- **Figure 3.3.2** is a Venn diagram that shows the union of two events as the total shaded area, with the intersection of the events being the overlapping region in the middle.
- If the events are **not disjoint**, then to find the probability of their **union** we take the sum of their individual probabilities and subtract the probability of their intersection (the part that was "counted twice").

3.3 Probability Rules

Additional Rules

- **Rule (4):** If two events E_1 and E_2 are **disjoint**, then

  $$\Pr\{E_1 \underline{\text{ or }} E_2\} = \Pr\{E_1\} + \Pr\{E_2\}$$

- **Rule (5):** For any two events E_1 and E_2 ,

  $$\Pr\{E_1 \underline{\text{ or }} E_2\} = \Pr\{E_1\} + \Pr\{E_2\} - \Pr\{E_1 \underline{\text{ and }} E_2\}$$

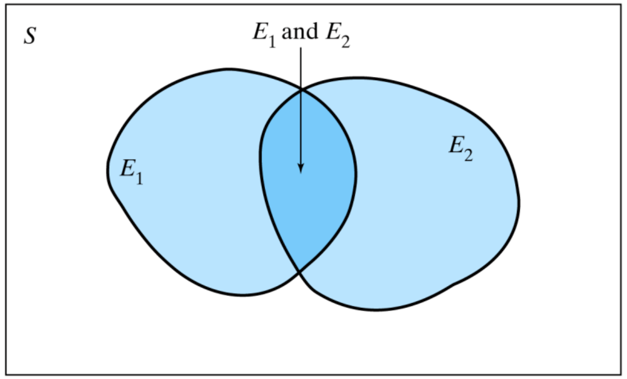


Figure 3.3.2 Venn diagram showing union (total shaded area) and intersection (middle area) of two events

Example 3.3.2 Hair color and eye color

- Pr{black hair or red hair}
- Pr{black hair or blue eyes}

Table 3.3.1 Hair color and eye color

		Hair color			
		Brown	Black	Red	Total
Eye color	Brown	400	300	20	720
	Blue	800	200	50	1,050
	Total	1,200	500	70	1,770

3.3 Probability Rules

Additional Rules

Example 3.3.2 Hair color and eye color (continued)

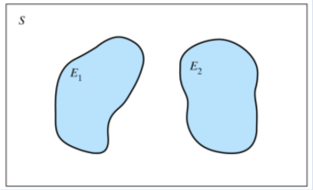
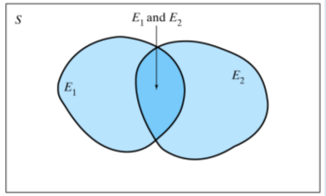


Table 3.3.1 Hair color and eye color

		Hair color			
		Brown	Black	Red	Total
Eye color	Brown	400	300	20	720
	Blue	800	200	50	1,050
	Total	1,200	500	70	1,770



- Pr{black hair or red hair}

  = Pr{black hair} + Pr{red hair}

  = 500/1770 + 70/1770

  = 570/1770

- Pr{black hair or blue eyes}

  = 500/1770 + 1050/1770 – 200/1770

  = 1350/1770

3.3 Probability Rules

Conditional probability

- Two events are said to be **independent** if knowing that one of them occurred does not change the probability of the other one occurring.
  - For example, if a coin is tossed twice, the outcome of the second toss is independent of the outcome of the first toss, since knowing whether the first toss resulted in heads or in tails does not change the probability of getting heads on the second toss.
- Events that are not independent are said to be **dependent**.

- When events are dependent, we need to consider the **conditional probability** of one event, given that the other event has happened.
- We use the notation **$Pr\{E_2 \mid E_1\}$** to represent the probability of E_2 happening, given that E_1 happened.

3.3 Probability Rules

Conditional probability

- **Conditional probability**: definition of the conditional probability of E_2, given E_1 happened:

  $\Pr\{E_2 \mid E_1\} = \Pr\{E_1 \ \underline{\text{and}}\ E_2\} / \Pr\{E_1\}$, provided that Pr{E 1 } > 0.

Multiplication Rules

- **Rule (6):** If two events E_1 and E_2 are **independent**,

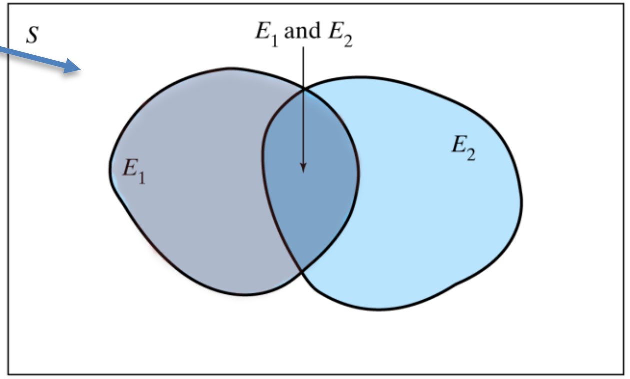
  $\Pr\{E_1 \ \underline{\text{and}}\ E_2\} = \Pr\{E_1\} \times \Pr\{E_2\}$

- **Rule (7):** For any two events E_1 and E_2 ,

  $\Pr\{E_1 \ \underline{\text{and}}\ E_2\} = \Pr\{E_1\} \times \Pr\{E_2 \mid E_1\}$

- **Rule (8):** <u>Rule of Total Probability</u>: For any two events E_1 and E_2 ,

  $\Pr\{E_1\} = \Pr\{E_1 \ \underline{\text{and}}\ E_2\} + \Pr\{E_1 \ \underline{\text{and}}\ E_2^C\} = \Pr\{E_2\} \times \Pr\{E_1 \mid E_2\} + \Pr\{E_2^C\} \times \Pr\{E_1 \mid E_2^C\}$



3.3 Probability Rules

Conditional probability

Example 3.3.3 Hair color and eye color

- Pr{blue eyes | black hair}

Table 3.3.1 Hair color and eye color

		Hair color			
		Brown	Black	Red	Total
Eye color	Brown	400	300	20	720
	Blue	800	200	50	1,050
	Total	1,200	500	70	1,770

Multiplication Rules

Example 3.3.5 Coin tossing

- Pr{heads twice}

Example 3.3.7 Hair color and eye color

- Pr{red hair and brown eyes}

3.3 Probability Rules

Conditional probability

Example 3.3.3 Hair color and eye color

- Pr{blue eyes|black hair}

  = Pr{black hair and blue eyes}/Pr{black hair}

  = (200/1770) / (500/1770) = 0.4

Table 3.3.1 Hair color and eye color

		Hair color			
		Brown	Black	Red	Total
Eye color	Brown	400	300	20	720
	Blue	800	200	50	1,050
	Total	1,200	500	70	1,770

Multiplication Rules

Example 3.3.5 Coin tossing

- Pr{heads twice} = Pr{heads on first toss} * Pr{heads on second toss} = 0.5 * 0.5 = 0.25

Example 3.3.7 Hair color and eye color

- Pr{red hair and brown eyes} = Pr{red hair} * Pr{brown eyes|red hair}
  = 70/1,770 x 20/70 = 20/1,770

3.3 Probability Rules

Multiplication Rules

- Rule of Total Probability: Rule (8) For any two events E_1 and E_2 ,

$$\Pr\{E_1\} = \Pr\{E_2\} \times \Pr\{E_1 \mid E_2\} + \Pr\{E_2^C\} \times \Pr\{E_1 \mid E_2^C\}$$

Example 3.3.8 Hand size

- Population - 60% female and 40% male.
- For a woman the probability of having a hand size smaller than 100 cm^2 is 0.31.
- For a man the probability of having a hand size smaller than 100 cm^2 is 0.08.

Table 3.3.2 Hand size

	Hand size		
	$< 100\ cm^2$	$\geq 100\ cm^2$	Total
Woman	186	414	600
Man	32	368	400
Total	218	782	1,000

- What is the probability that the randomly chosen person will have a hand size smaller than 100 cm^2 ?

3.3 Probability Rules

Multiplication Rules

- Rule of Total Probability: Rule (8) For any two events E_1 and E_2 ,

$$\Pr\{E_1\} = \Pr\{E_2\} \times \Pr\{E_1 \mid E_2\} + \Pr\{E_2^C\} \times \Pr\{E_1 \mid E_2^C\}$$

Example 3.3.8 Hand size

- What is the probability that the randomly chosen person will have a hand size smaller than 100 cm^2 ?
  - Pr{hand size <100}

    = Pr{woman} x Pr{hand size <100|woman}

    + Pr{man} x Pr{hand size <100|man}

    = 0.6 x 0.31 + 0.4 x 0.08

    = 0.218.

Table 3.3.2 Hand size

	Hand size		
	< 100 cm^2	≥ 100 cm^2	Total
Woman	186	414	600
Man	32	368	400
Total	218	782	1,000

3.4 Density Curves

Relative Frequency Histograms and Density Curves

- **Frequency distribution for a variable**: in Chapter 2 we discussed the use of a histogram to represent a frequency distribution for a variable.

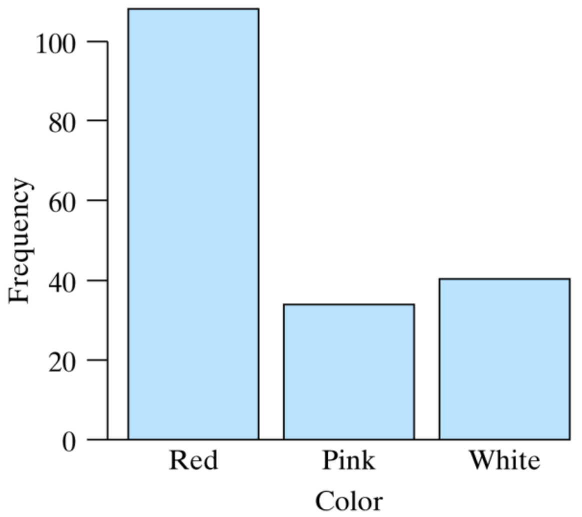
Example 2.2.1 Color of Poinsettias

- Poinsettias can be red, pink, or white. In one investigation of the hereditary mechanism controlling the color, 182 progeny of a certain parental cross were categorized by color.

- The bar graph in Figure 2.2.1 is a visual display of the results given in Table 2.2.1

Table 2.2.1 Color of 182 poinsettias

Color	Frequency (number of plants)
Red	108
Pink	34
White	40
Total	182



3.4 Density Curves

Relative Frequency Histograms and Density Curves

- **Relative frequency histogram**: histogram in which we indicate the proportion (i.e., the relative frequency) of observations in each category, rather than the count of observations in the category.

Example 3.4.1 Blood Glucose

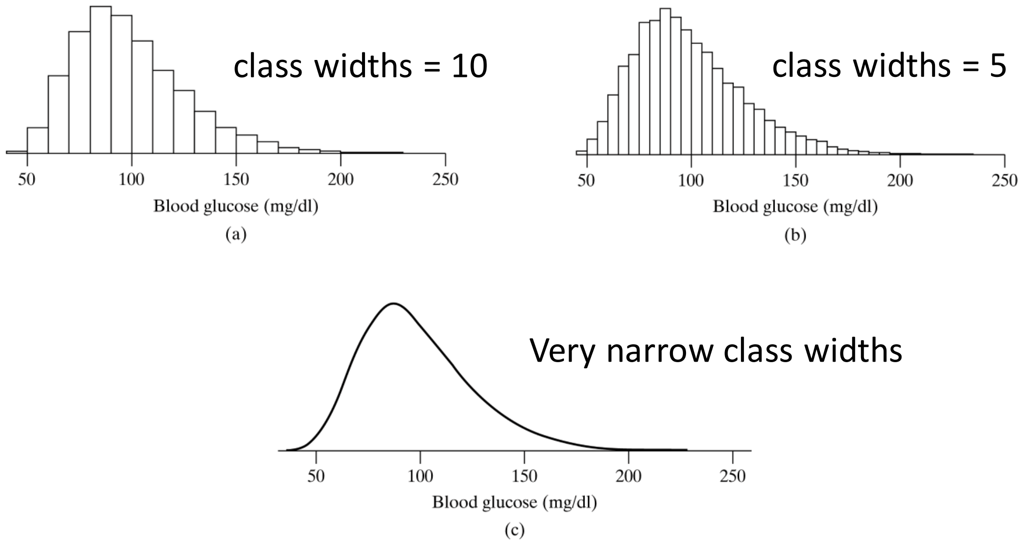


Figure 3.4.1 Different representations of the distribution of blood glucose levels in a population of women

- We can think of the relative frequency histogram as an approximation of the underlying true population distribution from which the data came.
- It is often desirable, especially when the observed variable is continuous, to describe a population frequency distribution by a smooth curve.
- We may visualize the curve as an idealization of a relative frequency histogram with very narrow classes.

3.4 Density Curves

Density Curves

- **Density curve**: a smooth curve representing a frequency distribution is called a density curve.

Interpretation of Density

- The **vertical coordinates** of a density curve are plotted on a scale called a **density scale**.
- The density curve is entirely above (or equal to) the x-axis and the area under the entire curve must be equal to 1 (Figure 3.4.3).
- When the density scale is used, relative frequencies are represented as areas under the curve. For any two numbers a and b,

  Area under density curve between a and b

  = Proportion of Y values between a and b

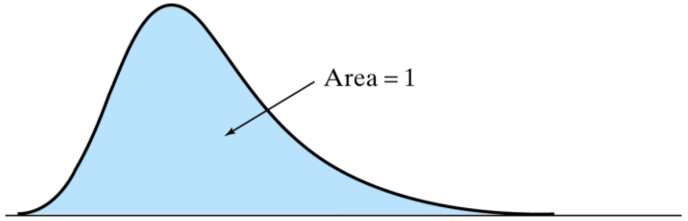


Figure 3.4.3 The area under an entire density curve must be 1

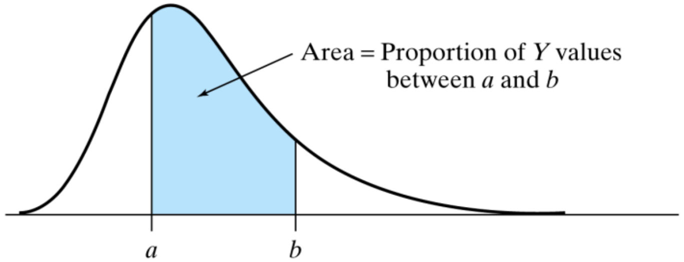


Figure 3.4.2 Interpretation of area under a density curve

3.4 Density Curves

The continuum paradox

- If we ask for the relative frequency of a single specific Y value, the answer is zero.
  - For example, suppose we want to determine from Figure 3.4.4 the relative frequency of blood glucose levels equal to 150. The area interpretation gives an answer of zero.
- If we are really asking for the relative frequency of glucose levels between 149.5 and 150.5 mg/dl, and the corresponding area is not zero.
- This is admittedly a paradoxical situation.
- In practice, the continuum paradox does not cause any trouble; **we simply do not discuss the relative frequency of a single Y value** (just as we do not discuss the length of a single point).

3.4 Density Curves

Relative Frequency Histograms and Density Curves

Example 3.4.2 Blood Glucose

- The shaded area of blood glucose distribution of Example 3.4.1 equal to 0.42.
- What is the meaning of 0.42?

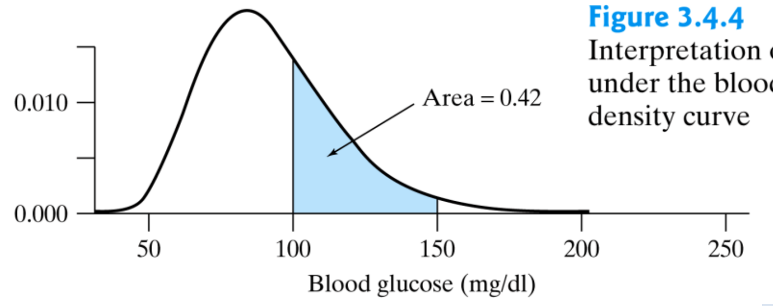


Figure 3.4.4 Interpretation of an area under the blood glucose density curve

- Pr{100 ≤ glucose level ≤ 150}
- Pr{100 < glucose level < 150}

3.4 Density Curves

Relative Frequency Histograms and Density Curves

Example 3.4.2 Blood Glucose

- The shaded area of blood glucose distribution of Example 3.4.1 equal to 0.42.
- What is the meaning of 0.42?
  - — about 42% of the glucose levels are between 100 mg/dl and 150 mg/dl

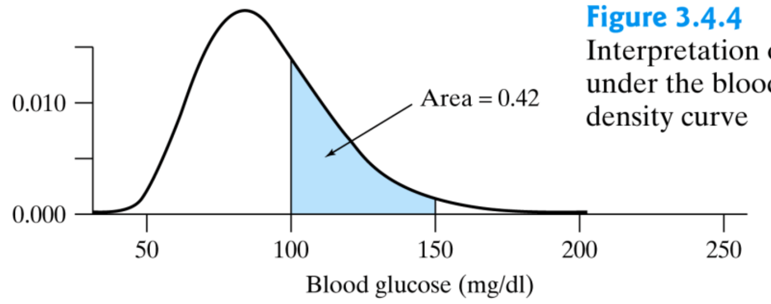


Figure 3.4.4 Interpretation of an area under the blood glucose density curve

- Pr{100 ≤ glucose level ≤ 150}
- Pr{100 < glucose level < 150}
  - — Pr{100 ≤ glucose level ≤ 150} = Pr{100 < glucose level < 150} = 0.42

- A probability for a continuous variable equals the area under the density curve for the variable between two points.

3.5 Random Variables

Random variable

- A **random variable** is simply a variable that takes on <u>numerical values</u> that depend on the outcome of a chance operation.

Example 3.5.1 Dice

- Consider the chance operation of tossing a die.
- What is the random variable?
- What is the probability of getting a 4?

3.5 Random Variables

Random variable

- A **random variable** is simply a variable that takes on <u>numerical values</u> that depend on the outcome of a chance operation.

Example 3.5.1 Dice

- Consider the chance operation of tossing a die.
- What is the random variable?
  - Let the **random variable** Y represent the <u>number of spots</u> showing.
  - The possible values of Y are Y = 1, 2, 3, 4, 5, or 6.
- What is the probability of getting a 4?
  - If the die is perfectly balanced so that each of the six faces is equally likely, then $Pr\{Y = 4\} = 1/6 \approx 0.17$

3.5 Random Variables

Mean and Variance of a Random Variable

- The **mean** of a discrete random variable Y is defined as

$$\mu_Y = \sum y_i \Pr(Y = yi)$$

where the y_i ’s are the values that the variable; the sum is taken over all possible values.

- The **mean** of a random variable is also known as the **expected value** and is often written as **E(Y)**; that is, **E(Y) =** μ_Y.

Example 3.5.6 Dice

- Consider rolling a die that is perfectly balanced.
- The random variable Y represent the number of spots showing.
- What is the expected value, or mean, of Y?

3.5 Random Variables

Mean and Variance of a Random Variable

Example 3.5.6 Dice

- Consider rolling a die that is perfectly balanced.
- The random variable Y represent the number of spots showing.
- What is the expected value, or mean, of Y?

E(Y) = $\mu_Y = \sum y_i \Pr(Y = yi)$

= 1x1/6 + 2x1/6 + 3x1/6 + 4x1/6 + 5x1/6 + 6x1/6

= 21/6

= 3.5

where the y_i ’s are the values that the variable; the sum is taken over all possible values.

3.5 Random Variables

Mean and Variance of a Random Variable

- The **variance** of a discrete random variable Y is defined as

$$\sigma^2{}_Y = \sum (y_i - \mu_Y)^2 \Pr(Y = yi)$$

where the y_i ’s are the values that the variable; the sum is taken over all possible values.

- We often write **VAR(Y)** to denote the variance of Y.
- The **standard deviation** of the random variable is $\boldsymbol{\sigma_Y}$: the square root of the variance.

Example 3.5.8 Dice

- What is the variance of Y?
- What is the standard deviation of Y?

3.5 Random Variables

Mean and Variance of a Random Variable

Example 3.5.6 Dice

- Consider rolling a die that is perfectly balanced.
- The random variable Y represent the number of spots showing.
- E(Y) = μ_Y =3.5
- What is the variance of Y?

$\sigma^2{}_Y = \sum(y_i - \mu_Y)^2 \Pr(Y = yi)$ = $(1\text{-}3.5)^2$ x Pr{Y=1} + $(2\text{-}3.5)^2$ x Pr{Y=2} + … + $(6\text{-}3.5)^2$ x Pr{Y=6} $\approx$ 2.9167

- What is the standard deviation of Y?

$\sigma_Y = \sqrt{2.9167} \approx 1.708$

- The preceding definitions are appropriate for discrete random variables.
- There are analogous definitions for continuous random variables, but they involve integral calculus and won't be presented here.

3.5 Random Variables

Adding and Subtracting Random Variables

- Rules for **Means** of Random Variables:
  - If Y is a random variable and a and b constants, then $\mu_{a+bY} = a + b\mu_Y$
- Rules for **Variances** of Random Variables:
  - If Y is a random variable and a and b constants, then $\sigma^2_{a+bY} = b^2\, \sigma^2_Y$

Example 3.5.9 Temperature

- The average summer temperature in a city is 81 ± 1 °F.
- Convert °F to °C, given: °C = (°F - 32) x (5/9).

3.5 Random Variables

Adding and Subtracting Random Variables

- Rules for **Means** of Random Variables:
  - If Y is a random variable and a and b constants, then $\mu_{a+bY} = a + b\mu_Y$
- Rules for **Variances** of Random Variables:
  - If Y is a random variable and a and b constants, then $\sigma^2_{a+bY} = b^2\,\sigma^2_Y$

Example 3.5.9 Temperature

- The average summer temperature in a city is 81 ± 1 °F.
- Convert °F to °C, given: °C = (°F - 32) x (5/9).
  - $\mu_{a+bY} = \sum(a + by_i)\Pr(Y = yi) = a + b\mu_Y = (81 - 32) \times (5/9) \approx 27.22$ °C
  - $\sigma^2_{a+bY} = \sum(a + byi - a - b\mu_Y)^2 \Pr(Y = yi) = b^2\,\sigma^2_Y = 1 \times (5/9)^2 \approx 0.3$ °C

3.6 The Binomial Distribution

Independent-Trials Model

- A series of **n** independent trials is conducted.
- Each trial results in “success” or “failure”.
- The probability of success is equal to the same quantity, **p**, for each trial, regardless of the outcomes of the other trails.

Example 3.2.7 Coin Tossing

- Give an example of Independent-Trials Model
- Tossing a coin twice
  - **n** = the number of tossing = 2
  - “Success”- head; “failure” – tail
  - p = ½

3.6 The Binomial Distribution

Binomial Random Variable

- A **binomial random variable** is a random variable that satisfies the following four conditions, abbreviated as **BInS**:
  - **Binary outcomes**: There are two possible outcomes for each trial (success and failure).
  - **Independent trials**: The outcomes of the trials are independent of each other.
  - **n** is fixed: The number of trials, n, is fixed in advance.
  - Same value of **p**: The probability of a success on a single trial is the same for all trials.

Binomial distribution

- **Binomial distribution** specifies the **probabilities** of various numbers of successes and failures when the basic chance operation consists of **n** independent trials.

3.6 The Binomial Distribution

What are the Binomial Random Variable and Binomial distribution in the examples?

- **Example 3.6.3 Albinism**

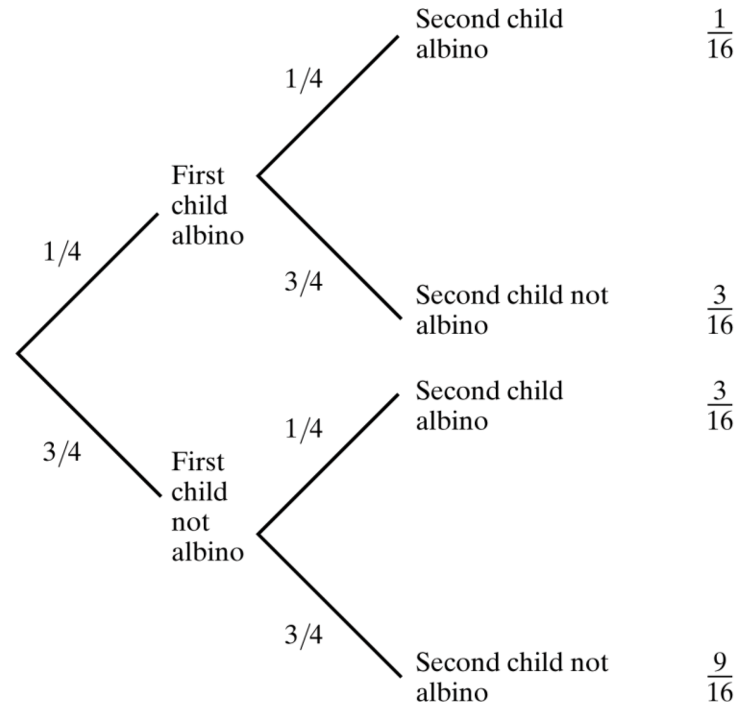


Figure 3.6.1 Probability tree for albinism among two children of carriers of the gene for albinism

- **Example 3.2.7 Coin Tossing**

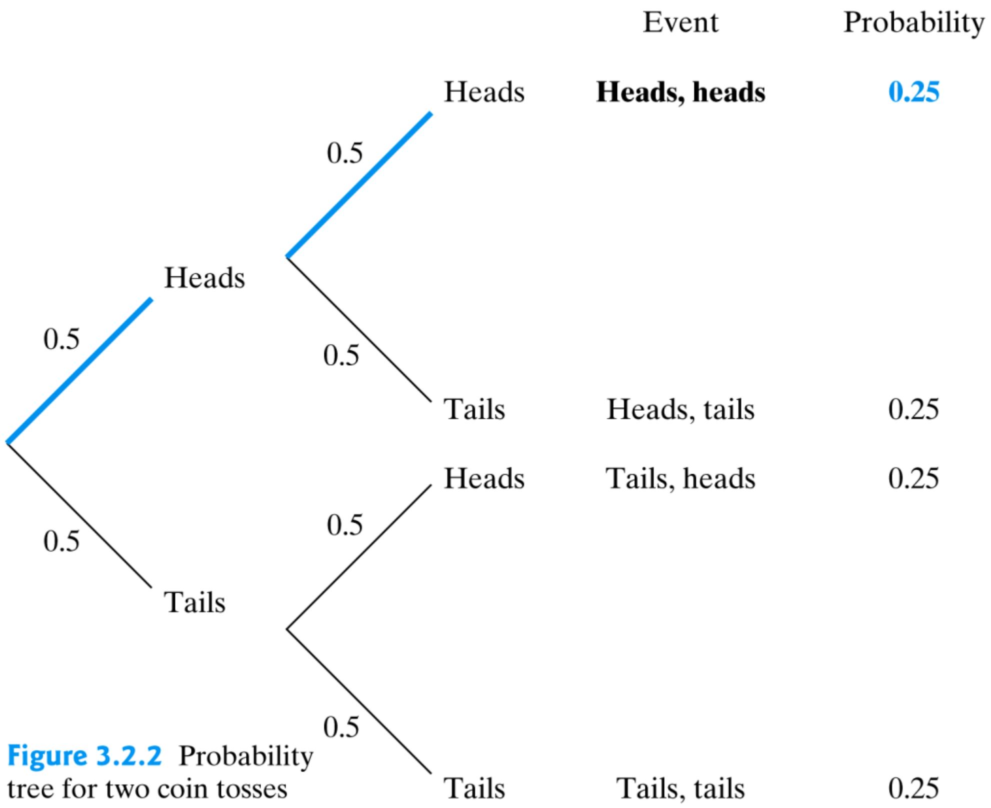
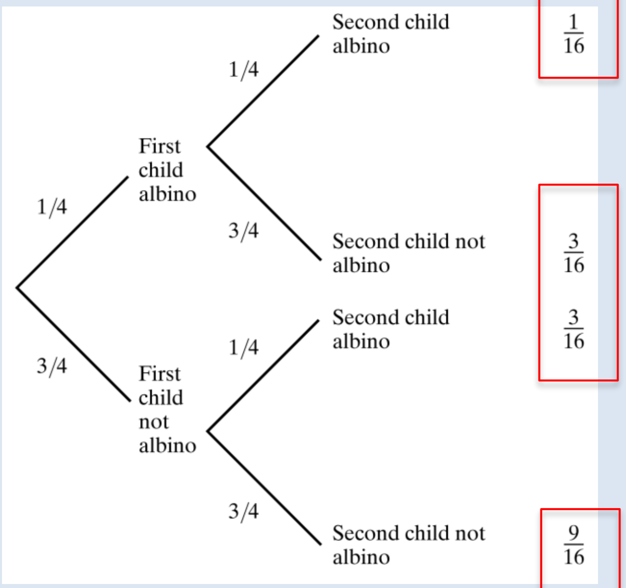


Figure 3.2.2 Probability tree for two coin tosses

3.6 The Binomial Distribution

What are the Binomial Random Variable and Binomial distribution in the examples?

- **Example 3.6.3 Albinism**
- Binomial random variable Y
  = number of albino child(ren) in the family



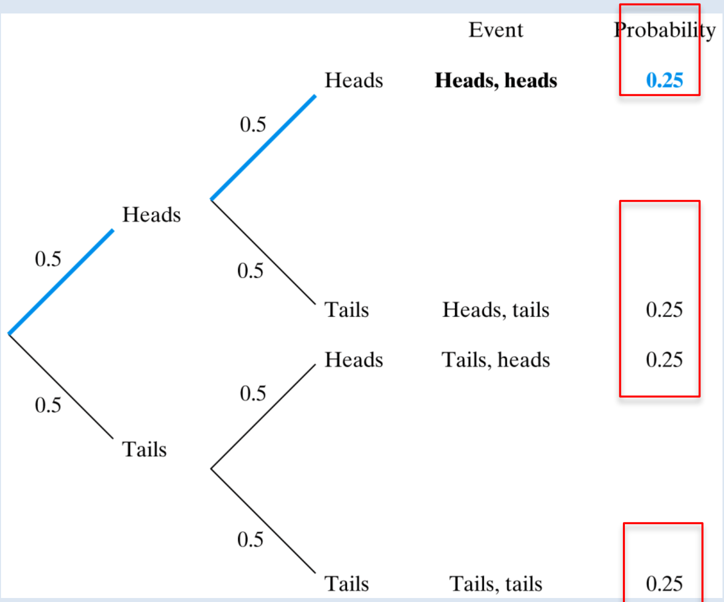
Y=2

Y=1

Y=0

binomial distribution with p =1/4 and n = 2.

- **Example 3.2.7 Coin Tossing**
- Binomial random variable Y
  = number of head after n toss



Y=2

Y=1

Y=0

binomial distribution with p =1/2 and n = 2.

3.6 The Binomial Distribution

The Binomial Distribution Formula

- For a binomial random variable Y, the probability that the **n** trials result in **j** successes (and n - j failures) is given by the following formula

$$\mathbf{Pr\{}\boldsymbol{j\ successes}\mathbf{\} = Pr(}\boldsymbol{Y = j}\mathbf{\} =} {}_{\boldsymbol{n}}\boldsymbol{C_j}\ \boldsymbol{p^j(1-p)^{n-j}}$$

Number of combination of having j success in n trials

Eg:
n =6
j = 4
${}_6C_4$ = 15

- The quantity ${}_{\boldsymbol{n}}\boldsymbol{C_j}$ appearing in the formula is called a **binomial coefficient**.
- Each binomial coefficient ${}_{\boldsymbol{n}}\boldsymbol{C_j}$ is an integer depending on n and on j.
- Values of binomial coefficients are given in **Statistical Tables** - Table 2** at the end of this book and can be found by the formula
- ${}_nC_j = \frac{n!}{j!(n-j)!}$, where x! (“x-factorial”) is defined for any positive integer x as

  x! = x(x - 1)(x - 2)…(2)(1) and 0! = 1.

3.6 The Binomial Distribution

The Binomial Distribution Formula

Example 3.6.6 and 3.6.7 blood type

- 85% of the population has Rh positive blood in US. Suppose we take a random sample of 6 persons and count the number with Rh positive blood.
- What is the probability that 4 persons (out of the 6 sampled) will have Rh positive blood?
- What is the probability that at least 4 persons (out of the 6 sampled) will have Rh positive blood?
- What is the probability that there is at least 1 person in the sample who has Rh negative blood.

3.6 The Binomial Distribution

The Binomial Distribution Formula

Example 3.6.6 and 3.6.7 blood type

- 85% of the population has Rh positive blood in US. Suppose we take a random sample of 6 persons and count the number with Rh positive blood.
  - "Binomial random variable Y: the number of persons, out of 6, with Rh positive blood."
  - "n=6, p=0.85"
- What is the probability that 4 persons (out of the 6 sampled) will have Rh positive blood?
  - $\Pr\{Y = 4\} = {}_nC_j p^j (1-p)^{n-j} = {}_6C_4 (0.85)^4 (1-0.85)^{6-4}$ = 0.1762
- What is the probability that at least 4 persons (out of the 6 sampled) will have Rh positive blood?
  - $\Pr\{Y \geq 4\} = \Pr\{Y = 4\} + \Pr\{Y = 5\} + \Pr\{Y = 6\}$ = 0.9526
- What is the probability that there is at least 1 person in the sample who has Rh negative blood.
  - $\Pr\{Y < 6\} = 1 - \Pr\{Y = 6\}$ = 0.6229

3.6 The Binomial Distribution

Mean and standard deviation of a Binomial

- For a binomial random variable, the **mean** (i.e., the average number of successes) is equal to $\boldsymbol{np}$.
  - This is an intuitive fact: The probability of success on each trial is p, so if we conduct n trials, then np is the expected number of successes.

- The **standard deviation** for a binomial random variable is given by

$$\sqrt{np\,(1-p)}$$

** A derivation of the result is given in Appendix 3.1 & 3.2.*

3.6 The Binomial Distribution

Mean and standard deviation of a Binomial

Example 3.2.7' Tossing Coin 10 times (continued)

- Binomial random variable Y: the number of head, out of n toss.
- n=10, p=0.5
- What is the mean and SD?

Example 3.6.8 blood type (continued)

- Binomial random variable Y: the number of persons, out of 6, with Rh positive blood.
- n=6, p=0.85
- What is the mean and SD?

3.6 The Binomial Distribution

Mean and standard deviation of a Binomial

Example 3.2.7’ Tossing Coin 10 times (continued)

- Binomial random variable Y: the number of head, out of n toss.
- n=10, p=0.5
- What is the mean and SD?
  - Mean: **mean = np** = 10 x 0.5 = 5. If we toss a fair coin 10 times, then we expect to get 5 heads, on average.
  - SD: Tossing a coin 10 times, the standard deviation SD = $\sqrt{10 \times 0.5 \times 0.5}$ = 1.58

Example 3.6.8 blood type (continued)

- Binomial random variable Y: the number of persons, out of 6, with Rh positive blood.
- n=6, p=0.85
- What is the mean and SD?
  - The average value of Y: **mean = np** = 6 x 0.85 = 5.1, which means that if we take many samples, each of size 6, and count the number of Rh positive persons in each sample, and then average those counts, we expect to get 5.1.
  - The standard deviation of Y: $\mathbf{SD} = \sqrt{\boldsymbol{np}\,(\mathbf{1}-\boldsymbol{p})}$ = $\sqrt{16 \times 0.85 \times 0.15}$ ≈ 0.87.

Summary

Chapter 3

- **3.1 Probability and the Life Sciences**
- **3.2 Introduction to Probability**
- **3.3 Probability Rules**
- **3.4 Density Curves**
- **3.5 Random Variables**
- **3.6 The Binomial Distribution**

Homework

Chapter 3

- **3.2.6 ;**
- **3.3.1 ; 3.3.2 ;**
- **3.4.2 ;**
- **3.5.1 ; 3.5.5 ; 3.5.6.**
- **3.6.1 ; 3.6.8 ;**