Introductory Statistics: A Problem-Solving Approach

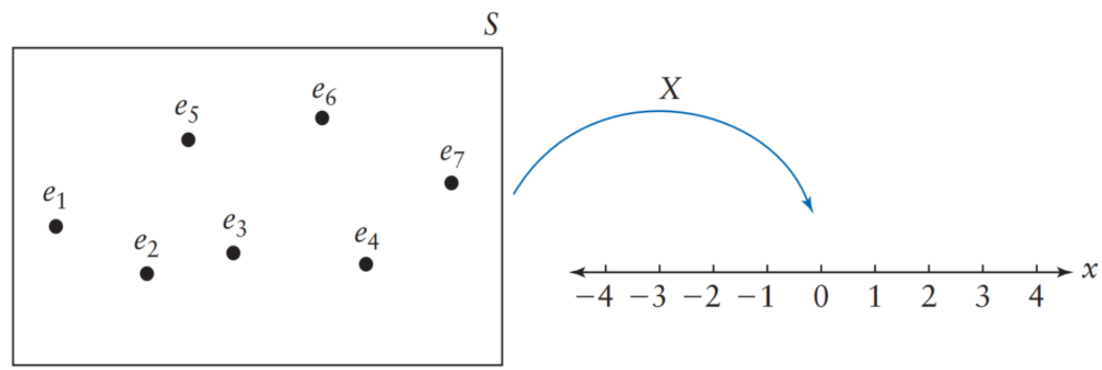
by Stephen Kokoska

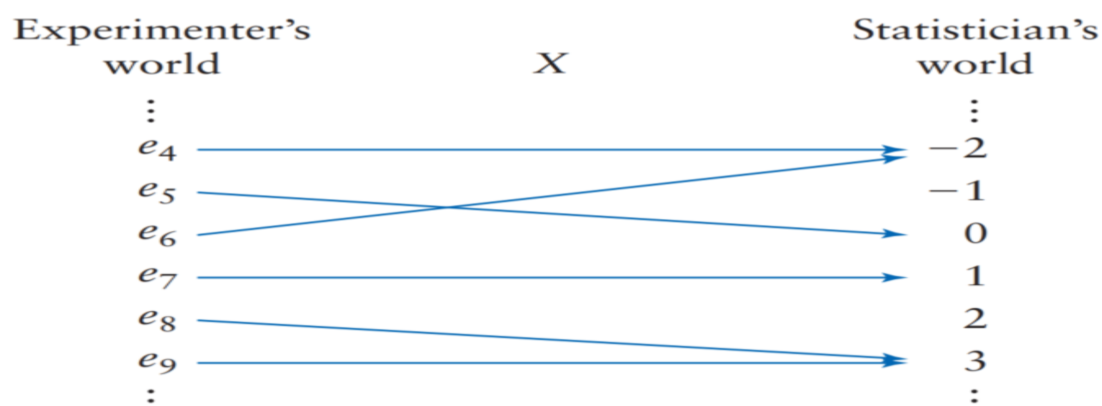
Chapter 5

Random Variables and Discrete Probability Distributions

5.1 Random Variable

- A **random variable** is a function that assigns a unique numerical value to each outcome in a sample space.
- These functions are called random variables because their values cannot be predicted with certainty before the experiment is performed.
- Capital letters, such as X and Y, are used to represent random variables.
- The rule for a random variable may be given by a formula, as a table, or even in words.
- The random variable X is the link between the experimental outcomes and the numerical values associated with each outcome.
- The visualization of a random variable illustrates the connection between each experimental outcome and the associated real number.





Types of Random Variables

- A random variable is **discrete** if the set of all possible values is finite, or countably infinite.
- A random variable is **continuous** if the set of all possible values is an interval of numbers.
- Discrete random variables are usually associated with counting, and continuous random variables are usually associated with measuring.
- Countable infinite means there are infinitely many possible values.
- The interval of possible values for a continuous random variable can be any interval, of any length, open or closed.
- In theory, a continuous random variable may assume any value in some interval (but not in reality).

Examples: Discrete or Continuous

Decide whether the associated random variable in each experiment is discrete or continuous:

(a) The Nassau Inn in Princeton, New Jersey, has 157 guest rooms. Let the random variable X be the number of occupied rooms on a randomly selected day. The possible values for X are finite: 0, 1, 2, …, 157. These values are distinct, disconnected points on a number line. X is discrete.

(b) One way to travel between parks at the Walt Disney World Resort is to use the monorail system. Let the random variable Y be the length of time it takes to travel from the Magic Kingdom to Epcot via monorail.
Y is a measurement, the time it takes to travel. The possible values for Y are any number in some interval. Y is continuous.

(c) An experiment consists of recording whether a randomly selected Ring video doorbell is good (G) or defective (B). S = {B, GB, GGB, GGGB, …}. Let the random variable X be the number of doorbells inspected until a defect is found. The values X can assume are 1, 2, 3, 4, …. The possible values are countably infinite, so X is discrete.

(d) Let the random variable Y be the length of the largest fish caught on the next party boat arriving back to the dock in Belmar, New Jersey.
Y is a measurement and can (theoretically) take on any value in some interval, say 5 to 25 inches. Y is continuous.

5.2 Probability Distributions for a Discrete Random Variable

The **probability distribution for a discrete random variable *X*** is a method for specifying all of the possible values of *X* and the probability associated with each value.

- ✓ A probability distribution for a discrete random variable may be presented in the form of an itemized listing, a table, a graph, or a function.
- ✓ A probability mass function (pmf), denoted as *p*, is the probability that a discrete random variable is equal to some specific value.

$$p(x) = \underbrace{P(X = x)}_{\text{Rule}}$$

Example: Suppose the random variable *Y* represents the number of wisdom teeth with which a person is born. The probabilities of *Y* taking on various values are as follows: 5/15 for 0 wisdom teeth; 4/15 for one; 3/15 for two; 2/15 for three; and 1/15 for all four. Represent the probability distribution for *Y* in multiple ways.

(1) A complete listing:

$P(Y = 0) = 5/15$
$P(Y = 1) = 4/15$
$P(Y = 2) = 3/15$
$P(Y = 3) = 2/15$
$P(Y = 4) = 1/15$

Example: Words of Wisdom (Teeth)

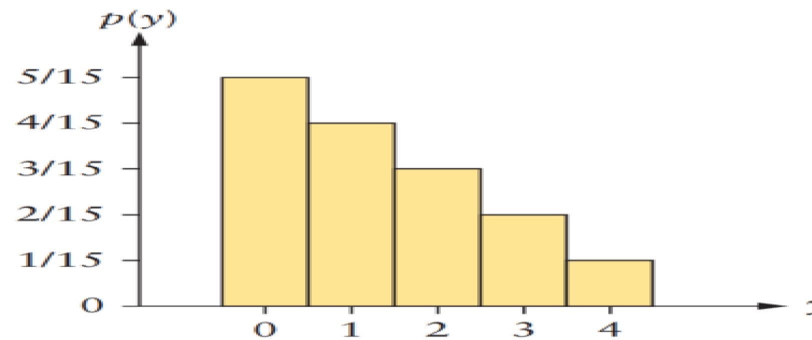
(2) A table of values and probabilities:

This kind of table is the most common way to present a probability distribution for a discrete random variable. It concisely lists all the values that *Y* can assume and the associated probabilities.

y	0	1	2	3	4
$p(y)$	5/15	4/15	3/15	2/15	1/15

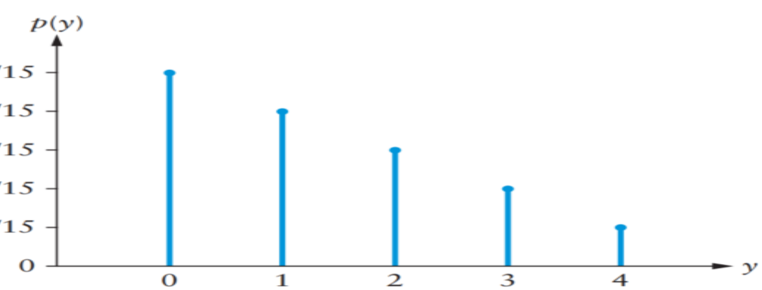
(3) A probability histogram:

The distribution of Y is represented graphically. A rectangle is drawn for each value y, centered at y, with height equal to p(y).



(4) A point representation:

Plot the points (y, $p(y)$) and draw a line from (y, 0) to (y, $p(y)$).



(5) A formula: $p(y) = \dfrac{5-y}{15}, \qquad y = 0,1,2,3,4$

This shows the rule for the probability mass function. To find $p(2)$, let $y = 2$ in the formula to find $p(2) = (5 - 2)/15 = 3/15$.

Properties of Probability Distributions

1. The probability that X takes on any value must be between 0 and 1.

$$0 \le p(x) \le 1$$

2. The sum of all the probabilities in a probability distribution for a discrete random variable must equal 1. $\sum_{\text{all } x} p(x) = 1$

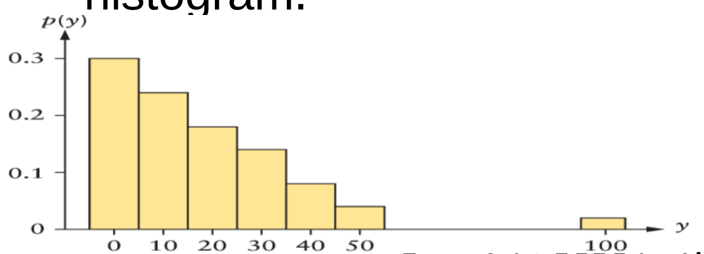
Example: The object in the Skeeball game is to score as many points as possible by landing the balls in holes with certain point values. The number of points earned on one roll is a random variable Y. Suppose Y has the following probability distribution:

y	0	10	20	30	40	50	100
$p(y)$	0.30	0.24	0.18	**?**	0.08	0.04	0.02

(a) Find $p(30)$

(b) Find $P(10 \le Y \le 40)$ and $P(10 < Y < 40)$.

(c) Construct the probability histogram.



$p(30) = 1 – [0.30 + 0.24 + 0.18 + 0.08 + 0.04 + 0.02]$
$= 1 – 0.86 = 0.14$

$P(10 \le Y \le 40) = p(10) + p(20) + p(30) + p(40)$
$= 0.24 + 0.18 + 0.14 + 0.08 = 0.64$

$P(10 < Y < 40) = p(20) + p(30) = 0.18 + 0.14 = 0.32$

5.3 Descriptive Measures for Discrete Random Variables

- A random variable may be used to model a population.
- The population descriptive measures of mean, variance, and standard deviation are inherent in and determined by the probability distribution.

Expected Value

Let X be a discrete random variable with probability mass function $p(x)$. The **mean**, or **expected value**, of X is

$$\underbrace{E(X) = \mu = \mu_X}_{\text{Notation}} = \underbrace{\sum_{\text{all } x} [x \cdot p(x)]}_{\text{Calculation}}$$

The mean of a random variable is a weighted average and is what happens on average.

Example: The First Columbia Bank & Trust Company offers four different fixed-rate home mortgages. The interest rate on each mortgage is related to the length of the loan. Suppose the length (in years) of a random mortgage is a discrete random variable X, with probability distribution given in the table. Find the mean of X.

x	10	15	20	30
$p(x)$	0.10	0.25	0.20	0.45

$\mu = (10)(0.10) + (15)(0.25) + (20)(0.20) + (30)(0.45)$

$= 1.00 + 3.75 + 4.00 + 13.50 = 22.5$

On average, the length of a mortgage is 22.25 years.

5.3 Descriptive Measures for Discrete Random Variables

Variance

Let X be a discrete random variable with probability mass function $p(x)$. The **variance** of X is

$$\underbrace{\mathrm{Var}(X) = \sigma^2 = \sigma_X^2}_{\text{Notation}} = \underbrace{\sum_{\text{all } x} [(x-\mu)^2 \cdot p(x)]}_{\text{Calculation}} = \underbrace{\mathrm{E}[(X-\mu)^2]}_{\text{Definition in terms of expected value}}$$

The variance is the expected value of the squared deviations about the mean

The standard deviation of X is the positive square root of the variance.

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

To calculate the variance using the definition:

(a) Find the mean μ of X.

(b) Find each difference: $(x - \mu)$.

(c) Square each difference: $(x - \mu)^2$.

(d) Multiply each squared difference by the associated probability.

(e) Sum the products.

In theory, the computational formula for the variance is faster and more accurate than using the definition:

$$\sigma^2 = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2 = \mathrm{E}(X^2) - \mu^2$$

$$\mathrm{E}(X^2) = \sum_{\text{all } x} [x^2 \cdot p(x)]$$

Example: Children in Day Care

Suppose the discrete random variable X, the age of a randomly selected child at the Precious Angels Child Care Center in Cleveland, Ohio, has the probability distribution given in the following table:

x	1	2	3	4	5	6	7
$p(x)$	0.05	0.10	0.15	0.25	0.20	0.15	0.10

a. Find the expected value, variance, and standard deviation of X.
b. Find the probability that the random variable takes on a value within 1 standard deviation of the mean.

$$\mathrm{E}(X) = \sum_{\text{all } x} [x \cdot p(x)]$$

If you select a child randomly from this center, his/her age is 4.30 years on average

μ=E(X)=(1)(0.05) + (2)(0.10) + (3)(0.15) + (4)(0.25) + (5)(0.20) + (6)(0.15) + (7)(0.10)

= 0.05 + 0.20 + 0.45 + 1.00 + 1.00 + 0.90 + 0.70 = 4.30

$$\mathrm{E}(X^2) = \sum_{\text{all } x} [x^2 \cdot p(x)]$$

$$\begin{aligned}\sigma^2 &= \mathrm{E}(X^2) - \mu^2 \\ &= 21.10 - (4.30)^2 \\ &= 21.10 - 18.49 \\ &= 2.61\end{aligned}$$

E(X^2)=(1^2)(0.05)+(2^2)(0.10)+(3^2)(0.15)+(4^2)(0.25)+(5^2)(0.20) +(6^2)(0.15 (7^2)(0.10)

= 0.05+0.40+1.35+4.00+5.00+5.40+4.90 = 21.10

$$\sigma = \sqrt{\sigma^2} = \sqrt{2.61} \approx 1.62$$

Example: Children in Day Care

(b) Find the probability that the random variable X takes on a value within 1 standard deviation of the mean.

$$\mu = 4.30$$
$$\sigma \approx 1.62$$

x	1	2	3	4	5	6	7
$p(x)$	0.05	0.10	0.15	0.25	0.20	0.15	0.10

$$\begin{aligned} P(\mu - \sigma \le X \le \mu + \sigma) &= P(4.30 - 1.62 \le X \le 4.30 + 1.62) \\ &= P(2.68 \le X \le 5.92) \\ &= P(X = 3) + P(X = 4) + P(X = 5) \\ &= 0.15 + 0.25 + 0.20 \\ &= 0.60 \end{aligned}$$

In the context of the problem, the probability that the age of a randomly selected child within 1 standard deviation of the mean is 0.60.

The Mean and Variance of Sums and Differences of Random Variables

Rules for Means

Rule 1: If X is a random variable and a and b are fixed numbers, then

$$\mu_{a+bX} = a + b\,\mu_X$$

Rule 2: If X and Y are random variables, then

$$\mu_{X+Y} = \mu_X + \mu_Y$$

Rule 3: If X and Y are random variables, then

$$\mu_{X-Y} = \mu_X - \mu_Y$$

Rules for Variances

Rule 1: If X is a random variable and a and b are fixed numbers, then

$$\sigma^2_{a+bX} = b^2\sigma^2_X$$

Rule 2: If X and Y are *independent* random variables, then

$$\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y$$

$$\sigma^2_{X-Y} = \sigma^2_X + \sigma^2_Y$$

Rule 3: If X and Y have correlation ρ, then

$$\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y + 2\,\rho\,\sigma_X\sigma_Y$$

$$\sigma^2_{X-Y} = \sigma^2_X + \sigma^2_Y - 2\,\rho\,\sigma_X\sigma_Y$$

Examples

Example1: The probability distribution of the number of courses taken in the fall and in the spring by students at a liberal arts college is given below:

Courses in the Fall	1	2	3	4	5	6
Probability	0.05	0.05	0.13	0.26	0.36	0.15

Courses in the Spring	1	2	3	4	5	6
Probability	0.06	0.08	0.15	0.25	0.34	0.12

For a randomly selected student, let X be the number of courses taken in the fall semester, and let Y be the number of courses taken in the spring semester. The means of these random variables are

$$\begin{aligned} \mu_X &= (1)(0.05) + (2)(0.05) + (3)(0.13) + (4)(0.26) + (5)(0.36) + (6)(0.15) \\ &= 4.28 \\ \mu_Y &= (1)(0.06) + (2)(0.08) + (3)(0.15) + (4)(0.25) + (5)(0.34) + (6)(0.12) \\ &= 4.09 \end{aligned}$$

The mean of the total number of courses taken for the academic year (fall and spring) is Z = X + Y

$$\begin{aligned} \mu_Z &= \mu_X + \mu_Y \\ &= 4.28 + 4.09 = 8.37 \end{aligned}$$

Examples

Example2: Consider a household where the monthly bill for natural-gas averages \$125 with a standard deviation of \$75, while the monthly bill for electricity averages \$174 with a standard deviation of \$41. The correlation between the two bills is −0.55.

Let X stands for the natural-gas bill and Y stands for the electricity bill. Then the variance of the total $X+Y$ is

$$\begin{aligned}\sigma^2_{X+Y} &= \sigma^2_X + \sigma^2_Y + 2\rho\sigma_X\sigma_Y \\ &= (75)^2 + (41)^2 + (2)(-0.55)(75)(41) \\ &= 3923.5\end{aligned}$$

$$\sigma_{X+Y} = \sqrt{3923.5} = 63$$

Example3: The following table gives the distribution of the number of servings of fruits and vegetables consumed per day in a population.

Number of servings X	0	1	2	3	4	5
Probability	0.3	0.1	0.1	0.2	0.2	0.1

What are the variance and standard deviation of the servings per week of fruits and vegetables in this population?

After finding $\mu_x = 2.2$ and $\sigma^2_x = 3.16$ $\quad \sigma^2_{7x} = 7^2\,\sigma^2_X = 49(3.16) = 154.84$ $\quad \sigma_{7x} \approx 12.44$

5.4 The Binomial Distribution

- The binomial random variable can be used to model many real-world populations.
- As with any random variable, there is a related experiment in the background. Consider the following experiments (and look for similarities):
  - ✓ Toss a coin 50 times and record the sequence of heads and tails.
  - ✓ Identify 40 volcanoes around the world and record whether each one erupts during the year.
  - ✓ Select a random sample of 25 customers at a fast-food restaurant and record whether each pays with exact change.
- All these experiments share four properties in common:
  1. The experiment consists of *n* identical trials.
  2. Each trial can result in only one of two possible (mutually exclusive) outcomes. One outcome is usually designated a success (*S*) and the other a failure (*F*). A success does not have to be a good thing.
  3. The outcomes of the trials are independent.
  4. The probability of a success *p* is constant from trial to trial.

5.4 The Binomial Distribution

- The **binomial random variable** maps each outcome in a binomial experiment to a real number. It is defined as the number of successes in *n* trials.
- The probability of a success is denoted by *p*. Therefore, P(*S*) = *p* and P(*F*) = 1 – *p* = *q*.
- A binomial random variable *X* is completely determined by the number of trials *n* and the probability of success *p*. Shorthand notation: *X* ~ B(*n*, *p*).
- **Suppose *X* is a binomial random variable with *n* trials and probability of a success *p*: *X* ~ B(*n*, *p*). Then**

$$p(x) = \mathrm{P}(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}, \; x = 0, 1, 2, \ldots, n$$

$$\binom{n}{x} = \text{Number of outcomes with } x \text{ successes}$$

$$p^x \cdot (1 - p)^{n-x} = \begin{matrix} \text{Probability of } x \text{ successes and } n - x \\ \text{failures in any single outcome} \end{matrix}$$

- If X is a binomial random variable with n trials and probability of a success p: X ~ B(n, p), then

$$\mu = np \qquad \sigma^2 = np(1 - p) \qquad \sigma = \sqrt{np(1 - p)}$$

Example: Digital Pizza

Online orders now make up more than half of all orders for some big chain pizza stores. According to a recent survey, approximately 60% of all pizza stores offer online ordering. Suppose 20 pizza orders are randomly selected.

(a) Find the probability that at most 12 are online orders.

(b) Find the probability that exactly 14 are online orders.

(c) Find the probability that at least 9 offer online orders.

(d) Find the mean and standard deviation of X variable

Let *X* be the number of pizza orders that are online orders. *X* ~ B(20, 0.60)

$$p(x) = \binom{20}{x} 0.6^{x} (1 - 0.6)^{20-x}, \ x = 0,1,2,\ldots,20$$

(a) P($X \leq 12$) = 0.5841 (Use technology or Table 1)

(b) P(X = 14) = P($X \leq 14$) – P($X \leq 13$)

= 0.8744 – 0.7500 = 0.1244

(c) P($X \geq 9$) = 1 - P(X < 9)

= 1- P(X ≤ 8) =1 – 0.0565 = 0.9435

(d) $\mu = np$ = (20)(0.6) = 12

$$\sigma = \sqrt{np(1-p)} = \sqrt{20(0.6)(0.4)} =$$

5.5 Other Discrete Distributions: Geometric, Poisson, Hypergeometric

Geometric Probability Distribution

- In a binomial experiment, n (the number of trials) is fixed, and the number of successes varies. The binomial random variable is the number of successes in n trials.
- In a geometric experiment, the number of successes is fixed at 1, and the number of trials varies. The geometric random variable is the number of trials necessary until the first success occurs. For example, when you continue to phone a friend until you get through. The number of calls necessary until the first success (reaching your friend) is the value of a geometric random variable.

Properties of a Geometric Experiment

1. The experiment consists of identical trials.
2. Each trial can result in only one of two possible outcomes: a success (S) or a failure (F).
3. The trials are independent.
4. The probability of a success p is constant from trial to trial.

Suppose *X* is a geometric random variable with probability of a success *p*: $X \sim G(p)$. Then $p(x) = P(X = x) = (1 - p)^{x-1} p \qquad x = 1, 2, \ldots$

$$\mu = \frac{1}{p} \qquad \sigma^2 = \frac{1-p}{p^2} \qquad \sigma = \frac{\sqrt{1-p}}{p}$$

Example: Bekins Men Are Careful, Quick, and Kind

According to the U.S. Census Bureau, approximately 11% of all people changed residences in 2016. Suppose researchers at the Bekins Moving Company randomly call people in the US and ask if they have moved in the last year.

(a) What is the probability that the fourth person called will be the first to have moved in the past year?

(b) What is the probability it will take at least six calls before the researchers speak to someone who has moved in the past year?

Let X be the number of calls necessary until the first mover is found. X is geometric with P(S) = 0.11 = p.

$$X \sim G(0.11)$$

$$p(x) = (1 - 0.11)^{x-1}(0.11)$$

(a) $P(X = 4) = (1 - 0.11)^{4-1}(0.11)$

$= (1 - 0.11)^{3}(0.11) = 0.0775$

(b) $P(X \geq 6) = 1 - P(X < 6)$

$= 1 - P(X \leq 5)$

$= 1 - [1 - (1 - 0.11)^{5}]$

$= 1 - 0.4416 = 0.5584$

If X is a geometric random variable with probability of success p, then $P(X \leq x) = 1 - (1 - p)^{x}$

$$\mu = \frac{1}{0.11} = 9.09 \qquad \sigma = \frac{\sqrt{1 - 0.11}}{0.11} \approx 8.58$$

Poisson Probability Distribution

A Poisson random variable is a count of the number of times the specific event occurs during a given unit of time, space, volume, distance, etc.

Examples:

- The number of arrivals to a hospital emergency room in a certain 30-min period.
- The number of asteroids that pass through Earth's orbit during a given year
- The number of bacteria in a milliliter of drinking water.

Properties of a Poisson Experiment

1. The probability that a specific event occurs in a given interval (of time, length, volume, etc.) is the same for all intervals.
2. The number of events that occur in any interval is independent of the number that occur in any other interval.

Suppose *X* is a Poisson random variable with parameter λ: ***X* ~ P(**λ**). Then**

$$p(x) = \mathrm{P}(X = x) = \frac{e^{-\lambda} \cdot \lambda^{x}}{x!}, \quad x = 0, 1, 2, 3, \ldots$$

e ≈ 2.71828 is the base of the natural logarithm.

$$\mu = \lambda \qquad \sigma^2 = \lambda \qquad \sigma = \sqrt{\lambda}$$

Example: Monthly Marine Occurrences

Suppose the mean number of fishing vessel accidents per month is six, and a random month is selected. Find the probability that

(a) there will be exactly four accidents involving fishing vessels.

(b) there will be at least eight accidents involving fishing vessels.

(c) the number of accidents involving fishing vessels will be within 1 standard deviation of the mean.

Let X be the number of fishing vessel accidents per month. X is Poisson with $\lambda = 6$.

$$X \sim P(6)$$

$$\mathrm{P}(X = x) = \frac{e^{-6} \cdot 6^x}{x!}, \quad x = 0, 1, 2, \ldots$$

(a) $\mathrm{P}(X = 4) = \dfrac{e^{-6} \cdot 6^4}{4!} = 0.1339$

(b) $\mathrm{P}(X \geq 8) = 1 - \mathrm{P}(X < 8)$

$= 1 - \mathrm{P}(X \leq 7)$ (Use technology or Table 2)

$= 1 - 0.7440 = 0.2560$

$$\mu = 6 \Rightarrow \sigma = \sqrt{6} = 2.4495$$

(c) $\mathrm{P}(\mu - \sigma \leq X \leq \mu + \sigma) = \mathrm{P}(6 - 2.4495 \leq X \leq 6 + 2.4495)$

$= \mathrm{P}(3.5505 \leq X \leq 8.4495)$

$= \mathrm{P}(4 \leq X \leq 8)$

$= \mathrm{P}(X \leq 8) - \mathrm{P}(X \leq 3)$ (Use technology or Table 2)

$= 0.8472 - 0.1512 = 0.6960$

Hypergeometric Probability Distribution

- The **hypergeometric probability distribution** arises from an experiment in which there is sampling without replacement from a finite population. The hypergeometric random variable is a count of the number of successes in a random sample of size *n*.
- For example, consider a shipment of 12 automobile tires, of which two are defective, and a random sample of four tires. A hypergeometric random variable may be defined as a count of the number of good tires selected.

Properties of a Hypergeometric Experiment

1. The population consists of *N* objects, of which *M* are successes and *N* – *M* are failures.
2. A sample of *n* objects is selected *without* replacement.
3. Each sample of size *n* is equally likely.

Suppose X is a hypergeometric random variable characterized by sample size n, population size N, and number of successes M:

X~ HG(*n*, *N*, *M*). Then

$$\mu = n\frac{M}{N}, \qquad \sigma^2 = \left(\frac{N-n}{N-1}\right)\cdot n\cdot\frac{M}{N}\cdot\left(1-\frac{M}{N}\right)$$

$$p(x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}, \qquad \max(0,\ n-N+M) \le x \le \min(n,\ M)$$

Example: Apple Keyboards

In May 2018, a class action lawsuit was filed in a northern California district court alleging that certain Apple keyboards in MacBook and MacBook Pro models were defective.
Suppose an Apple Store has 10 MacBook Pro models for sale. Two of the 10 MacBooks have defective keyboards and will fail soon after purchasers begin using them. Suppose four of the MacBooks are randomly selected. What is the probability that

(a) exactly two MacBooks will have good working keyboards?
(b) at least three MacBooks will have working keyboards?

Let X = number of working keyboards. X is hypergeometric with n = 4, N = 10, M = 8

$$p(x) = \frac{\binom{8}{x}\binom{10-8}{4-x}}{\binom{10}{4}}, \qquad \max(0,\ 4-\ 10+8) \le\ x \le \min(4,\ 8)$$

$$P(X = 2) = \frac{\binom{8}{2}\binom{10-8}{4-2}}{\binom{10}{4}} = \frac{\binom{8}{2}\binom{2}{2}}{\binom{10}{4}} = \frac{(28)(1)}{210} = 0.1333$$

$$P(X \ge 3) = P(X = 3) + P(X = 4)$$

$$= \frac{\binom{8}{3}\binom{2}{1}}{\binom{10}{4}} + \frac{\binom{8}{4}\binom{2}{0}}{\binom{10}{4}} = \frac{(56)(2)}{210} + \frac{(70)(1)}{210} = 0.8666$$