CHAPTER 3
Discrete Random Variables and Their Probability Distributions

Recap

- Why study the theory of probability ?

Recap

- Why study the theory of probability ?

- **Answer**: the probability of an observed event is needed to make *inferences* about a population.

Recap

- Why study the theory of probability ?
- **Answer**: the probability of an observed event is needed to make *inferences* about a population.
- What is probability distribution and why study them?

Recap

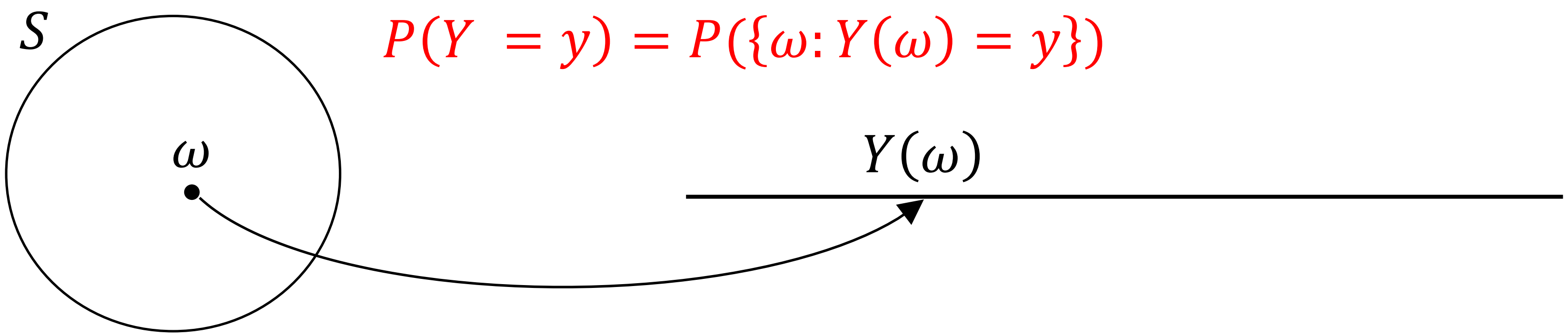
- Why study the theory of probability ?
- **Answer**: the probability of an observed event is needed to make *inferences* about a population.
- What is probability distribution and why study them?
- **Answer:**
    - The collection of probabilities is called the *probability distribution* of the random variable.
    - Knowledge of the probability distributions for random variables associated with *common types of experiments* will eliminate the need for solving the same probability problems over and over again.

Recap

DEFINITION 2.12

A *random variable* (r.v.) Y is a real-valued function $Y(\omega)$ over the sample space S of a random experiment, i.e., $Y : S \to \mathbb{R}$

$$P(Y = y) = P(\{\omega : Y(\omega) = y\})$$



- Randomness comes from ω ($Y(\omega)$ is a deterministic function)
- Notations:
  - ✓ Always use upper case letters for random variables ($X, Y, \dots$)
  - ✓ Always use lower case letters for values of random variables: $Y = y$ means that the random variable Y takes on the value y

Types of a Random Variable

- **Discrete**: X can assume only one of a countable number of values. Such r.v. can be specified by a *Probability Mass Function* (pmf). **(Chapter 3)**

- **Continuous**: X can assume one of a continuum of values and the probability of each value is 0. Such r.v. can be specified by a *Probability Density Function* (pdf). **(Chapter 4)**

- **Mixed**: X is neither discrete nor continuous. Such r.v. (as well as discrete and continuous r.v.s) can be specified by a *Cumulative Distribution Function* (cdf). **(Chapter 4)**

Discrete Random Variable

DEFINITION 3.1

A random variable Y is said to be *discrete* if it can assume only a finite or countably infinite number of distinct values.

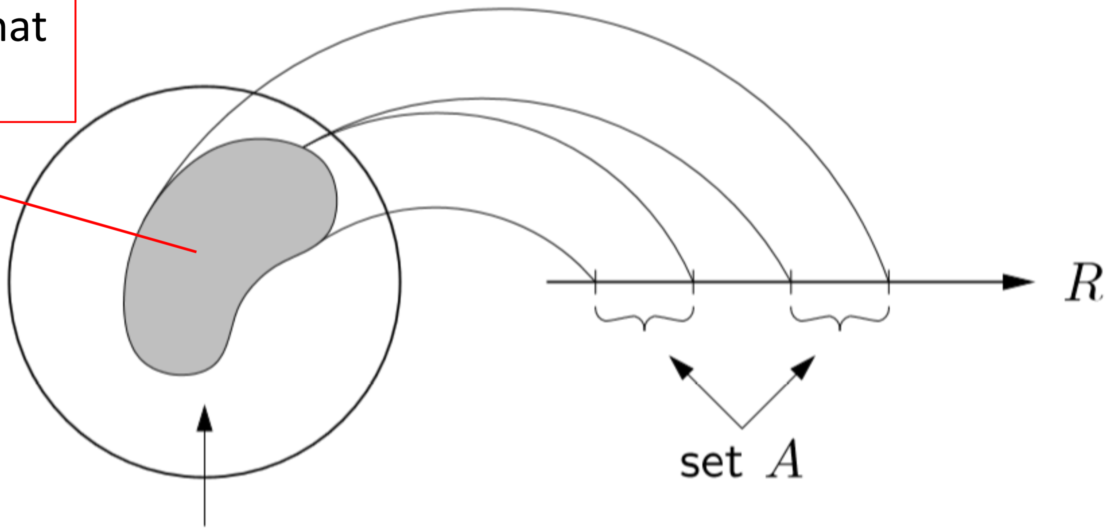
Examples:

- Y =The number of voters favoring a certain candidate or issue
- Y =The number of bacteria per unit area in the study of drug control

Specifying Random Variable

- Specifying a random variable means being able to determine the probability that $X \in A$ for any interval $A \subset \mathbb{R}$
- To do so, we consider the inverse image of the set A under $X(\omega)$, $\{w : X(\omega) \in A\}$

The subset of sample space that results in the event $X \in A$



inverse image of A under $X(\omega)$, i.e., $\{\omega : X(\omega) \in A\}$

- So, $X \in A$ iff $\omega \in \{w : X(\omega) \in A\}$, $P(\{X \in A\}) = P(\{w : X(\omega) \in A\})$

A is a set in $\mathbb{R}$ **Set in S**

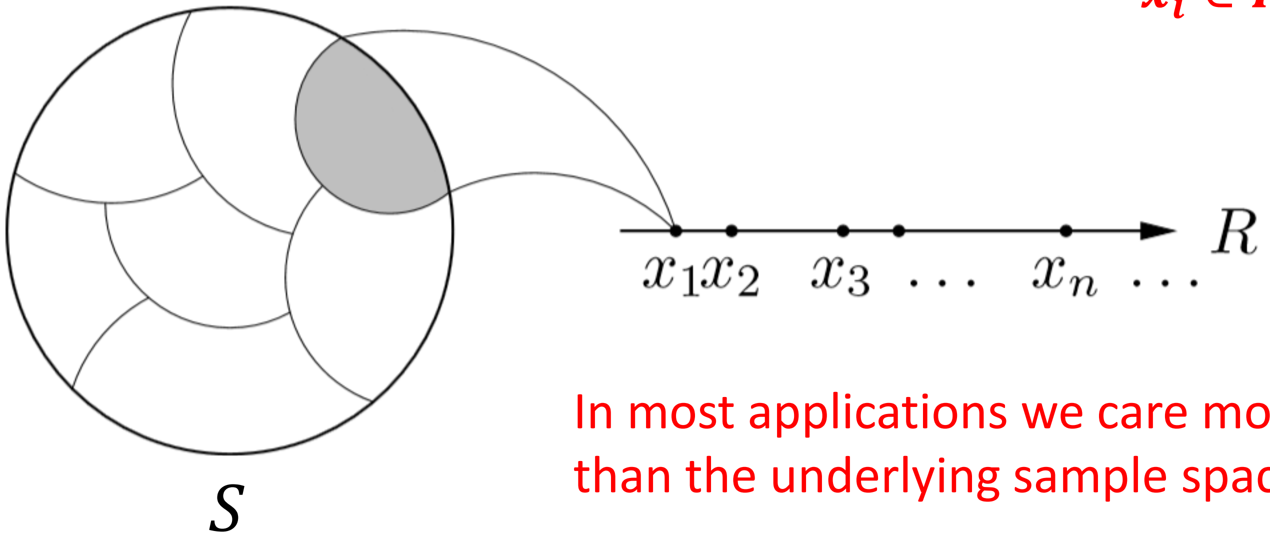
or in short

$$P\{X \in A\} = P\{w : X(\omega) \in A\}$$

Specifying Random Variable

- A random variable X is said to be **discrete** if for some countable set $\mathcal{X} \subset \mathbb{R}, i.e., \mathcal{X} = \{x_1, x_2, \ldots\}$, $P\{X \in \mathcal{X}\} = 1$
- Here $X(\omega)$ partitions S into the sets $\{\omega : X(\omega) = x_i\}$ for $i = 1, 2, \ldots$.
  - Therefore, to specify X, it suffices to know $P(\{X = x_i\})$ for all i

$\boldsymbol{x_i \in R}$



In most applications we care more about $X(\omega)$ (i.e., costs/measurements) than the underlying sample space S

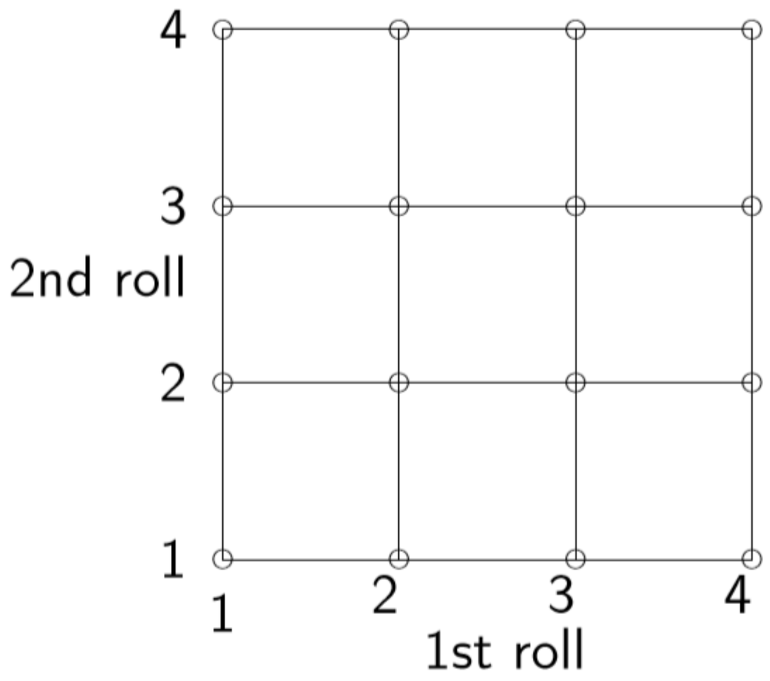
- A **discrete random variable** is thus completely specified by its *probability mass function* (pmf)

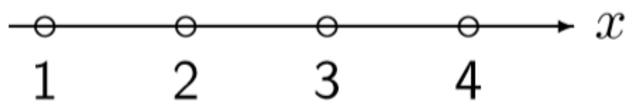
$$p_X(x) = P(\{X = x\}) \text{ for all } x \in \mathcal{X}$$

- Or we can write $p(x) = P(X = x)$

Specifying a Random Variable : Example

- Roll fair 4-sided die twice independently: Define the r.v. X to be the maximum of the two rolls. What is the $P(0.5 < X \leq 2)$?





Probability Distribution

DEFINITION 3.3

The *probability distribution* for a discrete variable Y can be represented by a formula, a table, or a graph that provides $p(y) = P(Y = y)$ for all y.

Probability Distribution : Example

EXAMPLE 3.1

A supervisor in a manufacturing plant has three men and three women working for him. He wants to choose two workers for a special job. Not wishing to show any biases in his selection, he decides to select the two workers at random. Let Y denote the number of women in his selection. Find the probability distribution for Y .

Probability Distribution : Example

SOLUTION 3.1

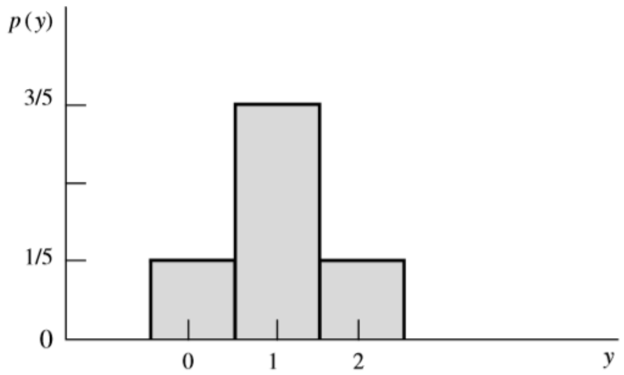
The supervisor can select two workers from six in $\binom{6}{2}$=15 ways. Hence, S contains 15 sample points, which we assume to be equally likely because random sampling was employed. Thus, $P(E_i)$ = 1/15, for $i = 1, 2, \ldots, 15$. The values for Y that have nonzero probability are 0, 1, and 2. The number of ways of selecting Y = 0 women is $\binom{3}{0}\binom{3}{2}$ because the supervisor must select zero workers from the three women and two from the three men. Thus, there are 3 sample points in the event $Y = 0$, and

$$p(0) = p(Y = 0) = \frac{\binom{3}{0}\binom{3}{2}}{15} = \frac{3}{15} = \frac{1}{5}$$

$$p(1) = p(Y = 1) = \frac{\binom{3}{1}\binom{3}{1}}{15} = \frac{9}{15} = \frac{3}{5}$$

$$p(2) = p(Y = 2) = \frac{\binom{3}{2}\binom{3}{0}}{15} = \frac{3}{15} = \frac{1}{5}$$

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



The formula for $p(y)$ also can be written in concise method as

$$p(y) = p(Y = y) = \frac{\binom{3}{y}\binom{3}{2-y}}{15}$$

discrete probability distribution

THEOREM 3.1

For any discrete probability distribution, the following must be true:

1. $0 \le p(y) \le 1$ for all y.

2. $\sum_y p(y) = 1$, where the summation is over all values of y with nonzero probability.

The Expected Value of a R.V.

DEFINITION 3.4

Let Y be a discrete random variable with the probability function $p(y)$. Then the *expected value* of Y, $E(Y)$, is defined to be

$$E(Y) = \sum_{y} yp(y)$$

- To be precise, the expected value of a discrete random variable is said to exist if the sum, as given earlier, is absolutely convergent – that is, if

$$E(Y) = \sum_{y} |y|p(y) < \infty$$

The Expected Value of a Function with R.V.

THEOREM 3.2

Let Y be a discrete random variable with probability function $p(y)$ and $g(Y)$ be a real-valued function of Y. Then the expected value of $g(Y)$ is given by $\mathrm{E}[g(Y)] = \sum_{\text{all } y} g(y)p(y)$.

- **Proof:**

The Expected Value of a Function with R.V.

THEOREM 3.2

Let Y be a discrete random variable with probability function $p(y)$ and $g(Y)$ be a real-valued function of Y. Then the expected value of $g(Y)$ is given by $\mathrm{E}[g(Y)] = \sum_{\text{all } y} g(y)p(y)$.

- **Proof:**

We prove the result in the case where the random variable Y takes on the finite number of values $y_1, y_2, \ldots, y_n$. Because the function $g(y)$ may not be one to-one, suppose that $g(Y)$ takes on values $g_1, g_2, \ldots, g_m$ (where $m \le n$). It follows that $g(Y)$ is a random variable such that for $i = 1, 2, \ldots, m$,

$$p[g(Y) = g_i] = \sum_{\substack{\text{all } y_j \text{ such that} \\ g(y_j)=g_i}} p(y_j) = p^*(g_i)$$

Thus, by Definition 3.4,

$$\begin{aligned} E[g(Y)] &= \sum_{i=1}^{m} g_i p^*(g_i) \\ &= \sum_{i=1}^{m} g_i \{\sum_{\{y_j: g(y_j)=g_i\}} p(y_j)\} \\ &= \sum_{i=1}^{m} \sum_{\substack{all\ y_j\ s.t. \\ g(y_j)=g_i}} g_i p(y_j) \\ &= \sum_{j=1}^{n} g(y_j)p(y_j). \end{aligned}$$

The Variance of R.V.

DEFINITION 3.5

If Y is a random variable with mean $E(Y) = \mu$, the variance of a random variable Y is defined to be the expected value of $(Y - \mu)^2$. That is, $V(Y) = E[(Y - \mu)^2]$. The *standard deviation* of Y is the positive square root of $V(Y)$.

- If $p(y)$ is an accurate characterization of the population frequency distribution (and to simplify notation, we will assume this to be true), then $E(Y) = \mu,\ V(Y) = \sigma^2$, the population variance, and σ is the population standard deviation.

Theorems

THEOREM 3.3

Let Y be a discrete random variable with probability function $p(y)$ and c be a constant. Then $E(c) = c$.

Proof:

- Consider the function $g(Y) = c$. By theorem 3.2,

$$E(c) = \Sigma_y cp(y) = c\Sigma_y p(y) = c(1) = c.$$

Theorems

THEOREM 3.4

Let Y be a discrete random variable with probability function $p(y)$, $g(Y)$ be a function of Y, and c be a constant. Then $E[cg(Y)] = cE[g(Y)]$.

Proof:

- By theorem 3.2,

$$E(cg(Y)) = \Sigma_y cg(y)p(y) = c\Sigma_y g(y)p(y) = cE[g(y)].$$

Theorems

THEOREM 3.5

Let Y be a discrete random variable with probability function $p(y)$ and $g_1(Y), g_2(Y), \ldots, g_k(Y)$ be k functions of Y. Then

$$E[g_1(Y) + g_2(Y) + \cdots + g_k(Y)] = E[g_1(Y)] + E[g_2(Y)] + \cdots + E[g_k(Y)].$$

Proof:

- We will demonstrate the proof only for the case $k = 2$, but analogous steps will hold for any finite k. By Theorem 3.2,

$$\begin{aligned} E[g_1(Y) + g_2(Y)] &= \Sigma_y [g_1(y) + g_2(y)]p(y) \\ &= \Sigma_y g_1(y)p(y) + \Sigma_y g_2(y)p(y) \\ &= E[g_1(Y)] + E[g_2(Y)]. \end{aligned}$$

Theorems

THEOREM 3.6

Let Y be a discrete random variable with probability function $p(y)$ and mean $E(Y) = \mu$; then

$$V(Y) = \sigma^2 = E[(Y - \mu)^2] = E(Y^2) - \mu^2.$$

Proof:

Theorems

THEOREM 3.6

Let Y be a discrete random variable with probability function $p(y)$ and mean $E(Y) = \mu$; then

$$V(Y) = \sigma^2 = E[(Y-\mu)^2] = E(Y^2) - \mu^2.$$

Proof:

- By Theorem 3.5,

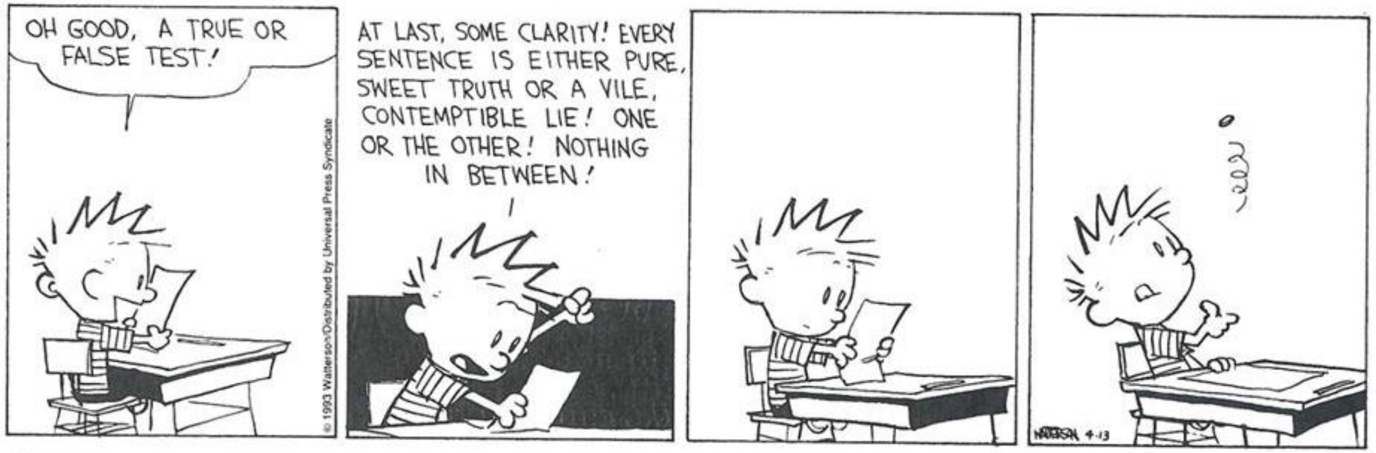
$$\begin{aligned}\sigma^2 = E[(Y-\mu)^2] &= E(Y^2 - 2\mu Y + \mu^2)\\ &= E(Y^2) - E(2\mu Y) + E(\mu^2)\end{aligned}$$

- Note that μ is a constant and $\mu = E(Y)$. Therefore,

$$\sigma^2 = E(Y^2) - 2\mu^2 + \mu^2 = E(Y^2) - \mu^2.$$

Motivation

- Some experiments consist of the observation of a sequence of *identical and independent* trials, each of which can result in *one of two outcomes*.
  - Each item leaving a manufacturing production line is either defective or non-defective.
  - Each shot in a sequence of firings at a target can result in a hit or a miss.
  - Each of n persons questioned prior to a local election either favors candidate Jones or not.



Binomial experiment

DEFINITION 3.6

A *binomial experiment* possesses the following properties:

1. The experiment consists of a fixed number, n, of identical trials.
2. Each trial results in one of two outcomes: success, S, or failure, F.
3. The probability of success on a single trial is equal to some value p and remains the same from trial to trial. The probability of a failure is equal to $q = (1 - p)$.
4. The trials are independent.
5. The random variable of interest is Y, the number of successes observed during the n trials.

- Determining whether a particular experiment is a binomial experiment requires examining the experiment for each of the characteristics just listed.
- It is important to realize that a success is not necessarily "good" in the everyday sense of the word.

Binomial Experiment : Example

EXAMPLE 3.6

Suppose that 40% of a large population of registered voters favor candidate Jones. A random sample of $n = 10$ voters will be selected, and Y, the number favoring Jones, is to be observed. Does this experiment meet the requirements of a binomial experiment?

Binomial Experiment : Example

SOLUTION 3.6

- The random variable of interest is then the number of successes in the ten trials.
- For the first person selected, the probability of favoring Jones (Success) is .4.
- The probability that the second person favors Jones is also .4 (Unconditional probability)
- However, the conditional probability of a success on later trials depends on the number of successes in the previous trials.
  - ✓ If the population of voters is large, removal of one person will not substantially change the fraction of voters favoring Jones, and the conditional probability that the second person favors Jones will be very close to .4.
- In general, **if the population is large** and **the sample size is relatively small**, the conditional probability of success on a later trial given the number of successes on the previous trials will stay approximately the same regardless of the outcomes on previous trials.
  - ➢ Thus, the trials will be approximately independent and so sampling problems of this type are approximately binomial.

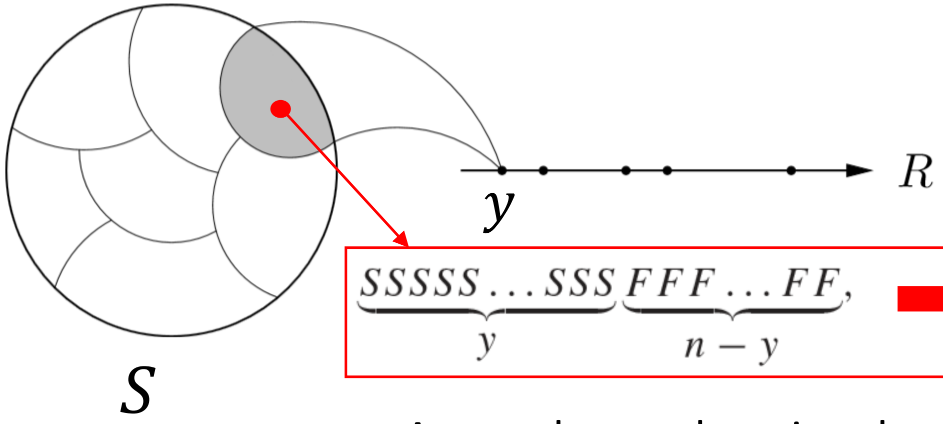
Binomial Distribution

DEFINITION 3.7

A random variable Y is said to have a *binomial distribution* based on n trials with success probability p if and only if

$$p(y) = \binom{n}{y} p^y q^{n-y}, \qquad y = 0, 1, 2, \ldots, n \ \ and \ \ 0 \le p \le 1$$

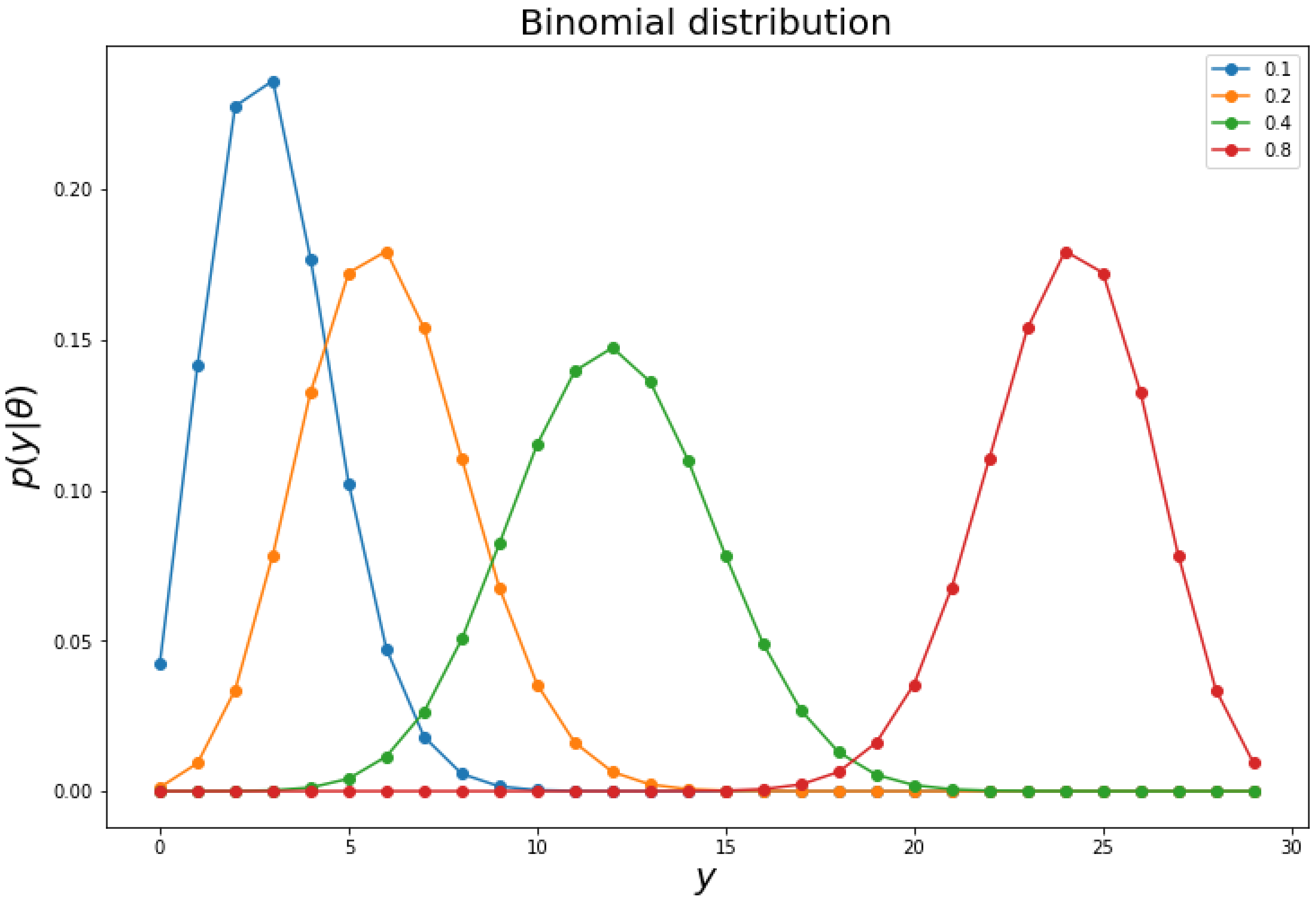
- The binomial probability distribution $p(y)$ can be derived by applying the sample point approach to find the probability that the experiment yields y successes



$$\underbrace{SSSSS\ldots SSS}_{y}\,\underbrace{FFF\ldots FF}_{n-y}, \;\Rightarrow\; \underbrace{ppppp\cdots ppp}_{y \text{ terms}}\,\underbrace{qqq\cdots qq}_{n-y \text{ terms}} = p^y q^{n-y}.$$

- Any such sample point also has probability $p^y q^{n-y}$
- The event $(Y = y)$ is made up of $\binom{n}{y}$ samples points, each with probability $p^y q^{n-y}$
- $p(y) = \binom{n}{y} p^y q^{n-y}$

Binomial Distribution



https://github.com/Jkparkaist/IE481/blob/master/Codes/L1_probabilityDistributions.ipynb

Binomial Expansion

- The term binomial experiment derives from the fact each trial results in one of two possible outcomes and that the probabilities $p(y), y = 0, 1, 2, \ldots, n$, are terms of the binomial expansion

$$(p+q)^n = \binom{n}{0} q^n + \binom{n}{1} p^1 q^{n-1}, \ldots, \binom{n}{y} p^y q^{n-y} +, \ldots, \binom{n}{n} p^n$$

$$= \quad p(0) \quad + \quad p(1), \ldots \quad , \qquad\qquad p(y) + , \ldots , \qquad\qquad p(n)$$

$$\sum_y p(y) = \sum_y \binom{n}{y} p^y q^{n-y} = (p+q)^n = 1^n = 1 \text{ because } p + q = 1$$

Binomial Distribution : Example

EXAMPLE 3.8

Experience has shown that 30% of all persons afflicted by a certain illness recover. A drug company has developed a new medication. Ten people with the illness were selected at random and received the medication; nine recovered shortly thereafter. What is the probability that at least nine of ten receiving the medication will recover?

Binomial Distribution : Example

SOLUTION 3.8

Let Y denote the number of people who recover.

$$P(Y = 9) = p(9) = \binom{10}{9}(.3)^9(.7) = .000138$$

$$P(Y = 10) = p(10) = \binom{10}{10}(.3)^{10}(.7)^0 = .000006$$

$$P(Y \geq 9) = p(9) + p(10) = .000138 + .000006 = .000144$$

The Mean and Variance of The Binomial Distribution

THEOREM 3.7

Let Y be a binomial random variable based on n trials and success probability p. Then

$$\mu = E(Y) = np \text{ and } \sigma^2 = V(Y) = npq.$$

Proof:

The Mean and Variance of The Binomial Distribution

THEOREM 3.7

Let Y be a binomial random variable based on n trials and success probability p. Then

$$\mu = E(Y) = np \text{ and } \sigma^2 = V(Y) = npq.$$

Proof:

- By Definitions 3.4, 3.7,

$$E(Y) = \sum_{y=1}^{n} yp(y) = \sum_{y=1}^{n} y\binom{n}{y}p^y q^{n-y}$$

$$= \sum_{y=1}^{n} y\frac{n!}{(n-y)!\,y!}p^y q^{n-y}$$

$$= np\sum_{y=1}^{n} \frac{(n-1)!}{(n-y)!\,(y-1)!}p^{y-1} q^{n-y}$$

$$= np\sum_{y=1}^{n} \frac{(n-1)!}{(n-1-z)!\,z!}p^{z} q^{n-1-z}$$

$$= np\sum_{z=0}^{n-1} \binom{n-1}{z}p^{z} q^{n-1-z} = np$$

$$\because \binom{n}{y} = \frac{n!}{(n-y)!\,y!}$$

Let
- $y - 1 = z$
- $n - y = n - 1 - z$

$$\sum_{z=0}^{n-1} \binom{n-1}{z}p^{z} q^{n-1-z} = \sum_{z=0}^{n-1} p(z) = 1$$

The Mean and Variance of The Binomial Distribution

THEOREM 3.7

Let Y be a binomial random variable based on n trials and success probability p. Then

$$\mu = E(Y) = np \text{ and } \sigma^2 = V(Y) = npq.$$

Proof:

- By Definitions 3.4, 3.7,

$$E[Y(Y-1)] = \sum_{y=1}^{n} y(y-1)p(y) = \sum_{y=2}^{n} y(y-1)\binom{n}{y}p^y q^{n-y}$$

$$= \sum_{y=2}^{n} y(y-1)\frac{n!}{(n-y)!\,y!}p^y q^{n-y}$$

$$= n(n-1)p^2 \sum_{y=1}^{n} \frac{(n-2)!}{(n-y)!\,(y-2)!}p^{y-2} q^{n-y}$$

$$= n(n-1)p^2 \sum_{y=1}^{n} \frac{(n-2)!}{(n-2-z)!\,z!}p^{z} q^{n-2-z}$$

$$= n(n-1)p^2 \sum_{z=0}^{n-2} \binom{n-2}{z}p^{z} q^{n-2-z} = n(n-1)p^2$$

$\because \binom{n}{y} = \frac{n!}{(n-y)!\,y!}$

Let

- $y-2=z$
- $n-y=n-2-z$

The Mean and Variance of The Binomial Distribution

THEOREM 3.7

Let Y be a binomial random variable based on n trials and success probability p. Then

$$\mu = E(Y) = np \text{ and } \sigma^2 = V(Y) = npq.$$

Proof:

- By Definitions 3.4, 3.7,

$$E[Y(Y-1)] = n(n-1)p^2$$

$$E[Y^2 - Y] = E[Y^2] - E[Y] = n(n-1)p^2$$

$$E[Y^2] = n(n-1)p^2 + E[Y] = n(n-1)p^2 + np$$

$$V(Y) = E[Y^2] - (E[Y])^2 = n(n-1)p^2 + np - (np)^2 = np - np^2 = np(1-p) = npq$$

Statistical Use of Binomial Distribution

EXAMPLE 3.10

Suppose that we survey 20 individuals working for a large company and ask each whether they favor implementation of a new policy regarding retirement funding. If, in our sample, 6 favored the new policy, find an estimate for p, the true but unknown proportion of employees that favor the new policy.

Statistical Use of Binomial Distribution

SOLUTION 3.10

If Y denotes the number among the 20 who favor the new policy, Y has a binomial distribution with $n = 20$ for some value of p. Thus,

$$P(Y = 6) = \binom{20}{6} p^6 (1 - p)^{14}.$$

We will use as our estimate for p the value that maximizes the probability of observing the value that we *actually observed* (6 in favor in 20 trials). How do we find the value of p that maximizes $P(Y = 6)$?

Because $\binom{20}{6}$ is a constant (relative to p) and $\ln(w)$ is an increasing function of w, the value of p that maximizes $P(Y = 6)$ is the same as the value of p that maximizes $\ln[p^6(1 - p)^{14}] = 6\ln(p) + 14\ln(1 - p)$.

If we take the derivative of $[6\ln(p) + 14\ln(1 - p)]$ with respect to p and set equals to 0, we obtain $p = 6/20$.

Because the second derivative of $[6\ln(p) + 14 \ln(1 - p)]$ is negative when $p = 6/20$, it follows that $P(Y = 6)$ is *maximized* when $p = 6/20$. Our estimate for p, based on 6 "successes" in 20 trials is therefore $6/20$.

As we will see in Chapter 9, the estimate that we just obtained is the *maximum likelihood* estimate for p and the procedure used above is an example of the application of the *method of maximum likelihood.*

Motivation

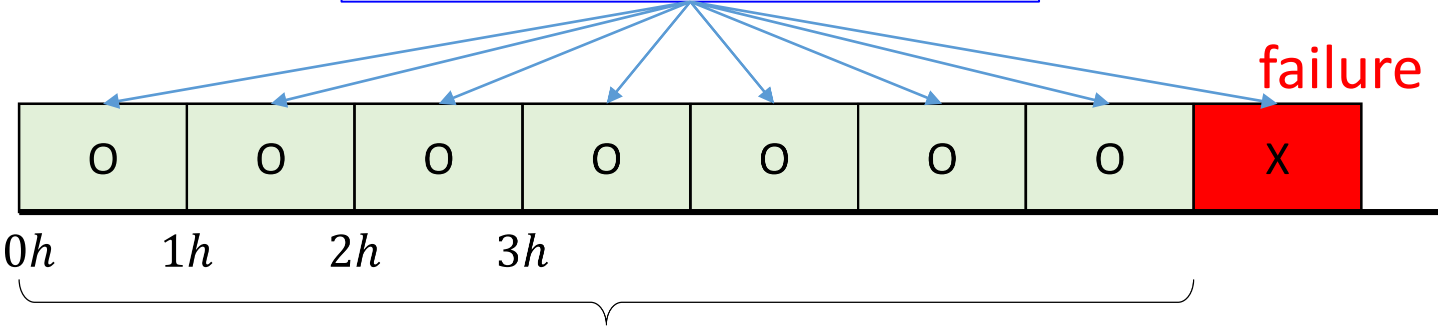
- The random variable with the geometric probability distribution is associated with an experiment that shares some of the characteristics of a binomial experiment.

- However the definition of random variable is different:
  - **Binomial**: $Y =$ *The number of success* among n trials
  - **Geometric**: $Y =$ The number of the trial on which the *first success* occurs

Motivation

- The geometric probability distribution is often used to model distributions of lengths of waiting times.
- For example, suppose that a commercial aircraft engine is serviced periodically so that its various parts are replaced at different points in time and hence are of varying ages.
- Then the probability of engine malfunction, p, during any randomly observed one-hour interval of operation might be the same as for any other one-hour interval.
- The length of time prior to engine malfunction is the number of one-hour intervals, Y, until the first malfunction. (For this application, engine malfunction in a given one-hour period is defined as a success.

$$\begin{cases} p \quad : \text{ Success (malfunction)} \\ 1-p: \text{ Normal} \end{cases}$$



$Y = 8 \qquad P(Y = 8) = ?$

Geometric Distribution

DEFINITION 3.8

A random variable Y is said to have a *geometric probability distribution* if and only if

$$p(y) = q^{y-1}p, \quad y = 1, 2, 3, \ldots, \quad 0 \le p \le 1.$$

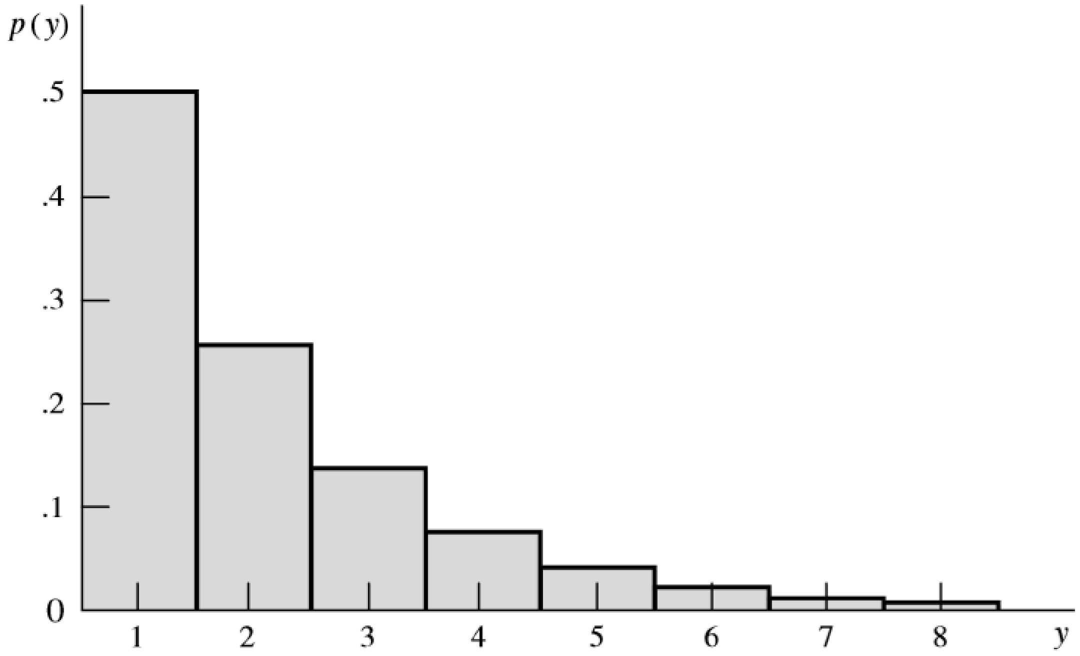
- The sample space S for the experiment contains the countably infinite set of sample points:

$$\begin{array}{lll}
E_1: & S & \text{(success on first trial)} \\
E_2: & FS & \text{(failure on first, success on second)} \\
E_3: & FFS & \text{(first success on the third trial)} \\
E_4: & FFFS & \text{(first success on the fourth trial)} \\
\vdots & & \\
E_k: & \underbrace{FFF\ldots F}_{k-1}S & \text{(first success on the } k^{\text{th}} \text{ trial)} \\
\vdots & &
\end{array}$$

- The numerical event $(Y = y)$ contains only E_y
- Because the trials are independent, for any $y = 1, 2, 3, \ldots,$

$$p(y) = P(Y = y) = P(E_y) = P(\underbrace{FFF \ldots F}_{y-1}S) = \underbrace{qqq \cdots q}_{y-1}p = q^{y-1}p.$$

Geometric Distribution



The geometric probability distribution, p=.5

Geometric Distribution : Example

EXAMPLE 3.11

Suppose that the probability of engine malfunction during any one-hour period is $p = .02$. Find the probability that a given engine will survive two hours.

Geometric Distribution : Example

SOLUTION 3.11

Let Y denote the number of one-hour intervals until the first malfunction.

$$P(survive\ two\ hours) = P(Y \geq 3) = \sum_{y=3}^{\infty} p(y)$$

Because $\sum_{y=1}^{\infty} p(y) = 1$,

$$P(survive\ two\ hours) = 1 - \sum_{y=1}^{2} p(y)$$

$$= 1 - p - qp = 1 - .02 - (.98)(.02) = .9604.$$

The Mean and Variance of Geometric Distribution

THEOREM 3.8

If Y is a random variable with a geometric distribution,

$$\mu = E(Y) = 1/p \text{ and } \sigma^2 = V(Y) = \frac{1-p}{p^2}$$

Proof:

The Mean and Variance of Geometric Distribution

THEOREM 3.8

If Y is a random variable with a geometric distribution,

$$\mu = E(Y) = 1/p \text{ and } \sigma^2 = V(Y) = \frac{1-p}{p^2}.$$

Proof:

$$E(Y) = \sum_{y=1}^{\infty} y q^{y-1} p$$

$$= p \sum_{y=1}^{\infty} y q^{y-1}$$

$$= p \frac{d}{dq}\left(\sum_{y=1}^{\infty} q^y\right)$$

$$= p \frac{d}{dq}\left(\frac{q}{1-q}\right) = p\left[\frac{1}{(1-q)^2}\right] = \frac{p}{p^2} = \frac{1}{p}$$

- larger values of p lead to higher probabilities for the smaller values of Y
- smaller values of p lead to lower probabilities for the larger values of Y
- Thus, the mean value of Y appears to be inversely proportional to p

The proposed approach is to express a series that cannot be summed directly as the derivative of a series for which the sum can be readily obtained.

The Mean and Variance of Geometric Distribution : Example

EXAMPLE 3.12

If the probability of engine malfunction during any one-hour period is $p = .02$ and Y denotes the number of one-hour intervals until the first malfunction, find the mean and standard deviation of Y.

The Mean and Variance of Geometric Distribution : Example

SOLUTION 3.12

Y has a geometric distribution with $p = .02$

Thus, $E(Y) = 1/p = 1/(.02) = 50$. We expect to wait 50 hours before encountering a malfunction.

Further, $V(Y) = .98/.0004 = 2450$, and $\sigma = \sqrt{2450} = 49.497$

The Mean and Variance of Geometric Distribution : Example

EXAMPLE 3.13

Suppose that we interview successive individuals working for the large company discussed in Example 3.10 and stop interviewing when we find the first person who likes the policy. If the fifth person interviewed is the first one who favors the new policy, find an estimate for p, the true but unknown proportion of employees who favor the new policy.

The Mean and Variance of Geometric Distribution : Example

SOLUTION 3.13

Let Y denote the number of individuals interviewed until we find the first person who likes the new retirement plan. Since Y has a geometric distribution with probability p, $P(Y = 5) = (1 - p)^4 p$.

To find the value of p that maximizes $P(Y = 5)$, we maximizes $\ln[(1 - p)^4 p] = 4\ln(1 - p) + \ln(p)$. If we take the derivative with respect to p and set this equal to 0, we obtain $p = 1/5$.

Because the second derivative of $[4\ln(1 - p) + \ln(p)]$ is negative when $p = 1/5$, it follows that $P(Y = 5)$ is *maximized* when $p = 1/5$. This is an example of the use of the *method of maximum likelihood* that will be studied in more detail in Chapter 9.

Motivation

- The geometric distribution handles the case where we are interested in the number of the trial on which the first success occurs.
- What if we are interested in knowing the number of the trial on which the second, third, or fourth success occurs? (geometric distribution only care about the first successes)
- The distribution that applies to the random variable Y equal to the number of the trial on which the r-th success occurs ($r = 2, 3, 4$, etc.) is the negative binomial distribution.

- What is the probability that he attempts Y shoots until achieving 4 success? ($r = 4$)

X	O	X	O	O	X	O	$y = 7$
O	O	X	O	O			$y = 5$
O	X	X	O	O	O		$y = 6$

Negative Binomial Probability Distribution

DEFINITION 3.9

A random variable Y is said to have a *negative binomial probability distribution* if and only if

$$p(y) = \binom{y-1}{r-1} p^r q^{y-r}, \qquad y = r, r+1, r+2, \ldots, \qquad 0 \le p \le 1.$$

- Let us select fixed values for r and y and consider events A and B, where
  - ✓ $A = \{\text{the first } (y-1) \text{ trials contain } (r-1) \text{ successes}\}$
    - ➢ $P(A) = \binom{y-1}{r-1} p^{r-1} q^{y-r}$
  - ✓ $B = \{\text{trial } y \text{ results in a success}\}.$
    - ➢ $P(B) = p$

$(\text{e.g.}, y = 5, r = 4)$

X	O	O	O	O
O	X	O	O	O
O	O	X	O	O
O	O	O	X	O

$P(A) = \binom{4}{3} p^3 q^1 \qquad P(B) = p$

Negative Binomial Probability Distribution

DEFINITION 3.9

A random variable Y is said to have a *negative binomial probability distribution* if and only if

$$p(y) = \binom{y-1}{r-1} p^r q^{y-r}, \qquad y = r, r+1, r+2, \ldots, \qquad 0 \le p \le 1.$$

- Let us select fixed values for r and y and consider events A and B, where
  - ✓ A = {the first $(y-1)$ trials contain $(r-1)$ successes}
    - ➢ $P(A) = \binom{y-1}{r-1} p^{r-1} q^{y-r}$
  - ✓ B = {trial y results in a success}.
    - ➢ $P(B) = p$
- Because we assume that the trials are independent, it follows that A and B are independent events,

$$p(y) = P(A \cap B) = P(A) \times P(B)$$

$$= \binom{y-1}{r-1} p^{r-1} q^{y-r}\, p = \binom{y-1}{r-1} p^r q^{y-r}$$

Mean and Variance of Negative Binomial Probability Distribution

THEOREM 3.9

If Y is a random variable with a negative binomial distribution,

$$\mu = E(Y) = rp \text{ and } \sigma^2 = V(Y) = \frac{r(1-p)}{p^2}.$$

Proof:

- These derivations will be much easier after we have developed some of the techniques of Chapter 5. For now, we state the following theorem without proof.

Negative Binomial Probability Distribution : Example

EXAMPLE 3.14

A geological study indicates that an exploratory oil well drilled in a particular region should strike oil with probability 0.2. Find the probability that the third oil strike comes on the fifth well drilled.

Negative Binomial Probability Distribution : Example

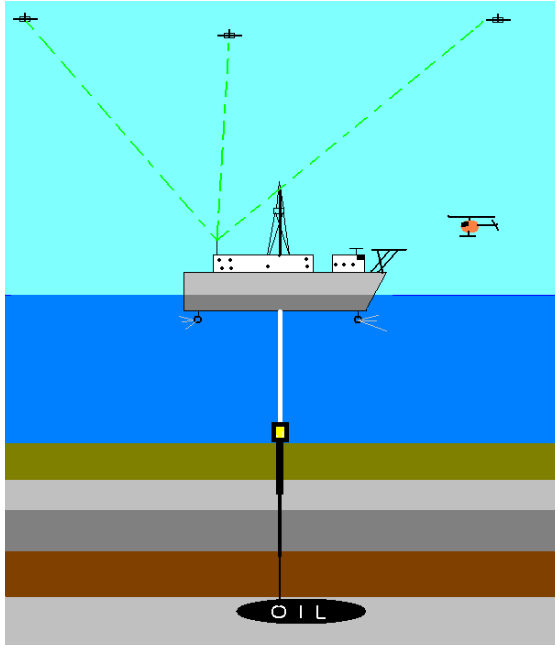
SOLUTION 3.14

Y has a negative binomial distribution with $p = .2$

We are interested in $r = 3$ and $y = 5$,

$$p(y) = \binom{y-1}{r-1} p^r q^{y-r}$$

$$P(Y = 5) = p(5) = \binom{4}{2}(.2)^3(.8)^2 = .0307$$



X	X	O	O	O
X	O	X	O	O
X	O	O	X	O
O	X	X	O	O
O	X	O	X	O
O	O	X	X	O

Negative Binomial Probability Distribution : Example

EXAMPLE 3.15

Alarge stockpile of used pumps contains 20% that are in need of repair. A maintenance worker is sent to the stockpile with three repair kits. She selects pumps at random and tests them one at a time. If the pump works, she sets it aside for future use. However, if the pump does not work, she uses one of her repair kits on it. Suppose that it takes 10 minutes to test a pump that is in working condition and 30 minutes to test and repair a pump that does not work. Find the mean and variance of the total time it takes the maintenance worker to use her three repair kits.

Negative Binomial Probability Distribution : Example

SOLUTION 3.15

- Let Y denote the number of the trial on which the third nonfunctioning pump is found.
- $r \ = 3$
- $P \ = 0.2$
- Y has a negative binomial distribution with $p = .2$

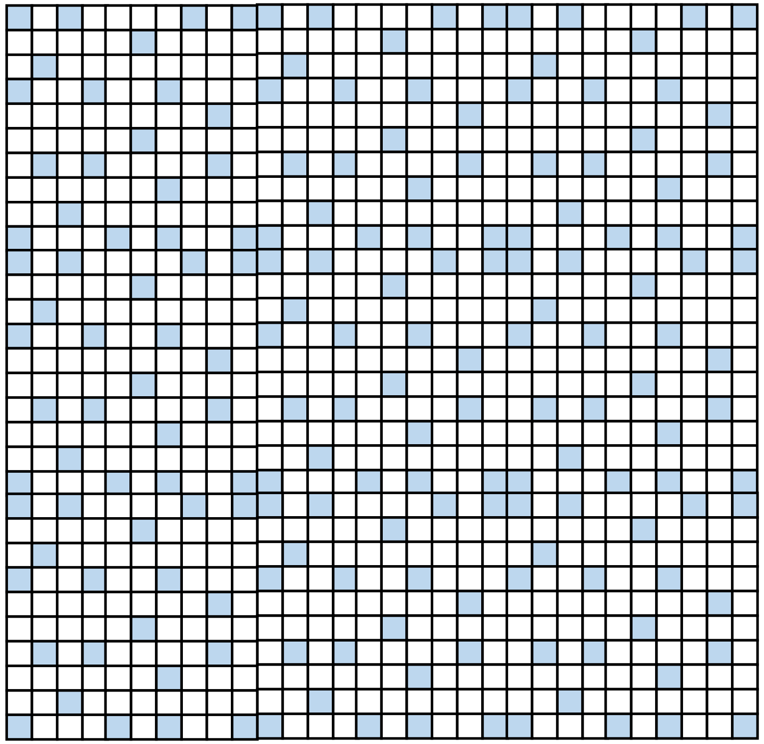
$$E(Y) = \frac{r}{p} = 15, V(Y) = \frac{r(1-p)}{p^2} = \frac{3(0.8)}{(0.2)^2} = 60$$

The total time to use the three kits is

$$T = 10Y + 3(20)$$
$$E(T) = 10E(Y) + 60 = 210$$
$$V(T) = 10^2 V(Y) = 100(60) = 6000$$
$$\sigma(T) = \sqrt{6000} = 77.46$$

Motivation

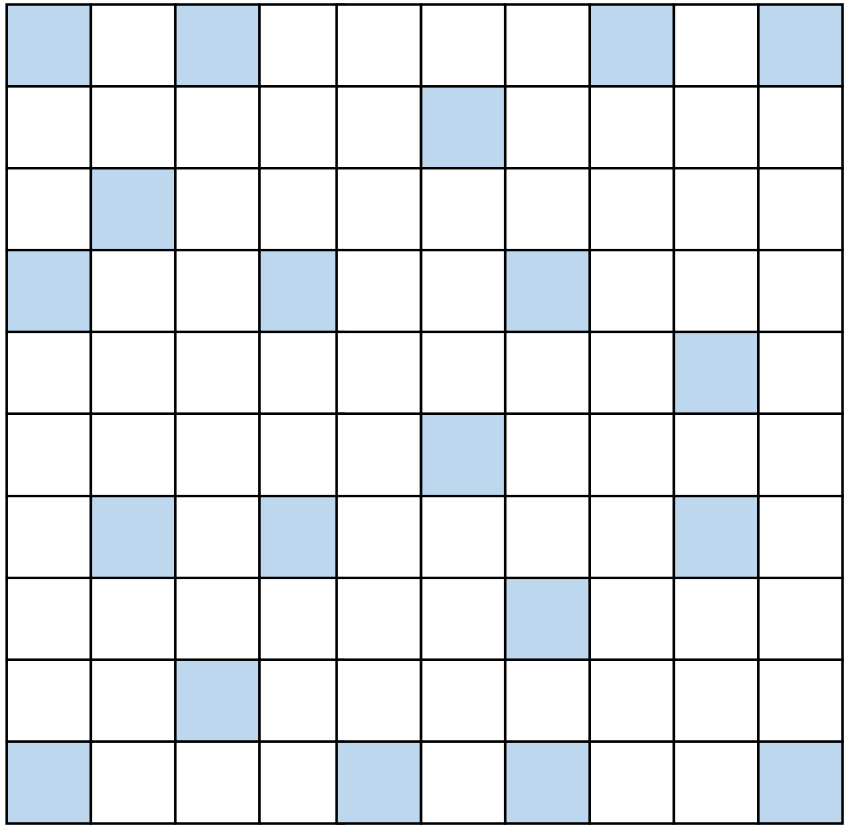
$N = 900, p = 0.2$



- $X_i = \begin{cases} 1 & \text{if } i\text{th voter favor JK} \\ 0 & \text{if } i\text{th voter does not favor JK} \end{cases}$
- $Y = \sum_{i=1}^{n} X_i$: the number voters favoring JK among n samples
- $P(X_1 = 1) = \frac{180}{900}$
- $P(X_2 = 1) = \frac{180}{900}$ (unconditional probability)
- $P(X_2 = 1 | X_1 = 1) = \frac{179}{899}$ (conditional probability)
- $P(X_2 = 1 | X_1 = 0) = \frac{180}{899}$ (conditional probability)
- $P(X_2 = 1) \approx P(X_2 = 1 | X_1 = 1) \approx P(X_2 = 1 | X_1 = 0)$

Motivation

$N = 100, p = 0.2$



- $X_i = \begin{cases} 1 & \text{if } i\text{th voter favor JK} \\ 0 & \text{if } i\text{th voter does not favor JK} \end{cases}$
- $Y = \sum_{i=1}^{n} X_i$: the number voters favoring JK among n samples

- $P(X_1 = 1) = \frac{20}{100}$
- $P(X_2 = 1) = \frac{20}{100}$ (unconditional probability)
- $P(X_2 = 1|X_1 = 1) = \frac{19}{99}$ (conditional probability)
- $P(X_2 = 1|X_1 = 0) = \frac{20}{99}$ (conditional probability)
- $P(X_2 = 1) \neq P(X_2 = 1|X_1 = 1) \neq P(X_2 = 1|X_1 = 0)$

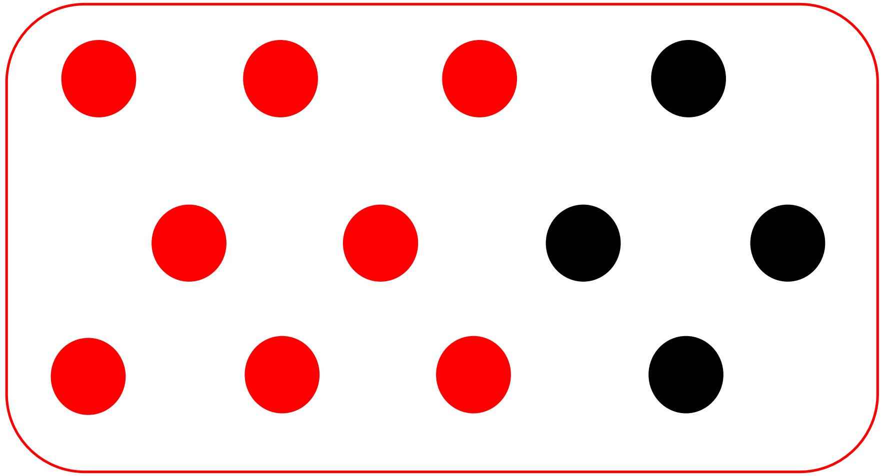
Motivation

- If the sample size n is small relative to the population size N, the distribution of Y **could** be approximated by a binomial distribution.
- If n is large relative to N, the conditional probability of selecting a supporter of Jones on a later draw would be significantly affected by the observed preferences of persons selected on earlier draws.
  - Thus the trials were not independent and the probability distribution for Y **could not** be approximated adequately by a binomial probability distribution.

Motivation

- Suppose that a population contains a finite number N of elements that possess one of two characteristics.
- Thus, r of the elements might be red and $b = N - r$, black.
- A sample of n elements is randomly selected from the population, and
- the random variable of interest is Y, the number of red elements in the sample.
- This random variable has what is known as the hypergeometric probability distribution

$$N = 12, n = 5, r = 8$$



$$p(Y = 3) = \frac{\binom{8}{3}\binom{12-8}{5-3}}{\binom{12}{5}}$$

Hypergeometric Probability Distribution

DEFINITION 3.10

A random variable Y is said to have a *hypergeometric probability distribution* if and only if

$$p(y) = \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}},$$

where y is an integer $0, 1, 2, \ldots, n$, subject to the restrictions $y \leq r$ and $n - y \leq N - r$.

- N and r are fixed
- We select n items among N items

Hypergeometric Probability Distribution

THEOREM 3.10

If Y is a random variable with a hypergeometric distribution,

$$\mu = E(Y) = \frac{nr}{N} \text{ and } \sigma^2 = V(Y) = n\left(\frac{r}{N}\right)\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right).$$

- Deriving closed form expressions for the resulting summations is somewhat tedious.
- In Chapter 5 we will develop methods that permit a much simpler derivation

Observations:

- If we define $p = \frac{r}{N}$ and $q = 1 - p = \frac{N-r}{N}$, then we can rewrite the mean and variance:

$$\mu = np \text{ and } \sigma^2 = npq\left(\frac{N-n}{N-1}\right)$$

- As $N \to \infty$, $\frac{N-n}{N-1} \to 1$. Thus, $\sigma^2 \to npq$ (same with Binomial distribution)

Hypergeometric Probability Distribution : Example

EXAMPLE 3.17

An industrial product is shipped in lots of 20. Testing to determine whether an item is defective is costly, and hence the manufacturer samples his production rather than using a 100% inspection plan. A sampling plan, constructed to minimize the number of defectives shipped to customers, calls for sampling five items from each lot and rejecting the lot if more than one defective is observed. (If the lot is rejected, each item in it is later tested.)

If a lot contains four defectives, what is the probability that it will be rejected?
What is the expected number of defectives in the sample of size 5?
What is the variance of the number of defectives in the sample of size 5?

Hypergeometric Probability Distribution : Example

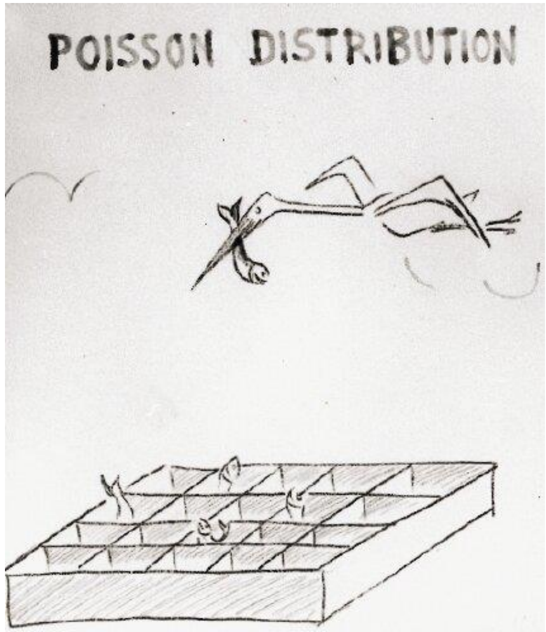
SOLUTION 3.17

Let Y equal the number of defectives in the sample. Then $N = 20, r = 4$, and $n = 5$.
The lot will be rejected if $Y = 2, 3$, or 4. Then

$$P(\text{rejecting the lot}) = P(Y \geq 2) = p(2) + p(3) + p(4) = 1 - p(0) - p(1)$$
$$= 1 - \frac{\binom{4}{0}\binom{16}{5}}{\binom{20}{5}} - \frac{\binom{4}{1}\binom{16}{4}}{\binom{20}{5}} = 1 - .2817 - .4696 = .2487$$

The mean and variance of the number of defectives in the sample of size 5 are
$\mu = \frac{(5)(4)}{20} = 1$ and $\sigma^2 = 5\left(\frac{4}{20}\right)\left(\frac{20-4}{20}\right)\left(\frac{20-5}{20-1}\right) = .632$

Motivation



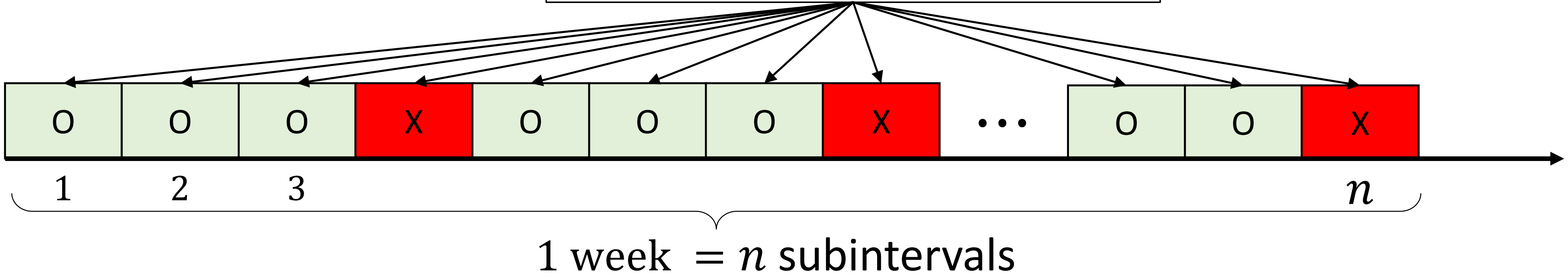
- The Poisson probability distribution often provides a good model for the probability distribution of the number Y of rare events that occur in space, time, volume, or any other dimension, where λ is the average value of Y.

- It provides a good model for the probability distribution of:
  - the number of automobile accidents given a unit of time
  - the number of industrial accidents given a unit of time
  - the number of other types of accidents in a given unit of time.
  - the number of telephone calls handled by a switchboard in a time interval
  - the number of radioactive particles that decay in a particular time period
  - the number of errors a typist makes in typing a page,
  - the number of automobiles using a freeway access ramp in a ten-minute interval.

Motivation

- Suppose that we want to find the probability distribution of the number of automobile accidents at a particular intersection during a time period of one week.
- Think of the time period, one week in this example, as being split up into n subintervals, *each of which is so small that at most one accident could occur in it with probability different from zero*.
- Denoting the probability of an accident in any subinterval by p, we have, for all practical purposes,

$$\begin{cases} p & : \text{Success (accident)} \\ 1-p & : \text{no accident} \\ 0 & : \text{more than one accident} \end{cases}$$



- If the occurrence of accidents can be regarded as independent from interval to interval, $Y =$ the total number of accidents has a binomial distribution with n trials.

Poisson Probability Distribution

DEFINITION 3.11

A random variable Y is said to have a *Poisson probability distribution* if and only if

$$p(y) = \frac{\lambda^y}{y!}e^{-\lambda}, \qquad y = 0, 1, 2, \ldots, \qquad \lambda > 0.$$

$$\begin{aligned}
\lim_{n\to\infty}\binom{n}{y}p^y(1-p)^{n-y} &= \lim_{n\to\infty}\frac{n(n-1)\cdots(n-y+1)}{y!}\left(\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{n-y} \\
&= \lim_{n\to\infty}\frac{\lambda^n}{y!}\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-y}\frac{n(n-1)\cdots(n-y+1)}{n^y} \\
&= \frac{\lambda^n}{y!}\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-y}\left(\frac{n}{n}\right)\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\times\cdots\times\left(1-\frac{y-1}{n}\right) \\
&= \frac{\lambda^n}{y!}e^{-\lambda}
\end{aligned}$$

$$\because \lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda}$$

- Because the binomial probability function converges to the Poisson, the Poisson probabilities can be used to approximate their binomial counterparts for large n, small p, and $\lambda = np$ less than, roughly, 7.

Poisson Probability Distribution : Example

EXAMPLE 3.18

Show that the probabilities assigned by the Poisson probability distribution satisfy the requirements that $0 \leq p(y) \leq 1$ for all y and $\sum_y p(y) = 1$

Poisson Probability Distribution : Example

SOLUTION 3.18

- Because $\lambda > 0$, it is obvious that $p(y) > 0$ for $y = 0, 1, 2, \ldots$, and that $p(y) = 0$ otherwise.

$$\sum_{y=0}^{\infty} p(y) = \sum_{y=0}^{\infty} \frac{\lambda^y}{y!} e^{-\lambda} = e^{-\lambda} \sum_{y=0}^{\infty} \frac{\lambda^y}{y!} = e^{-\lambda} e^{\lambda} = 1$$

Because $\sum_{y=0}^{\infty} \frac{\lambda^y}{y!}$ is a series expansion of e^{λ}

The Mean and Variance of Poisson Probability Distribution

THEOREM 3.11

If Y is a random variable possessing a Poisson distribution with parameter λ, then

$$\mu = E(Y) = \lambda \text{ and } \sigma^2 = V(Y) = \lambda.$$

Proof:

The Mean and Variance of Poisson Probability Distribution

THEOREM 3.11

If Y is a random variable possessing a Poisson distribution with parameter λ, then

$$\mu = E(Y) = \lambda \text{ and } \sigma^2 = V(Y) = \lambda.$$

Proof:

$$E(Y) = \sum_{y} yp(y) = \sum_{y=0}^{\infty} y \frac{\lambda^y}{y!} e^{-\lambda}$$

$$= \sum_{y=1}^{\infty} \frac{\lambda^y e^{-\lambda}}{(y-1)!}$$

$$= \lambda \sum_{y=1}^{\infty} \frac{\lambda^{y-1} e^{-\lambda}}{(y-1)!}$$

$$= \lambda \sum_{z=0}^{\infty} \frac{\lambda^z e^{-\lambda}}{z!} = \lambda \sum_{z=0}^{\infty} p(z) = \lambda \qquad \text{(set } z = y - 1\text{)}$$

Poisson Probability Distribution : Example

EXAMPLE 3.22

Industrial accidents occur according to a Poisson process with an average of three accidents per month. During the last two months, ten accidents occurred. Does this number seem highly improbable if the mean number of accidents per month, μ, is still equal to 3? Does it indicate an increase in the mean number of accidents per month?

Poisson Probability Distribution : Example

SOLUTION 3.22

The number of accidents in *two* months, Y, has a Poisson probability distribution with mean $\lambda = 2(3) = 6$. The probability that Y is as large as 10 is $P(Y \geq 10) = \frac{\sum_{y=10}^{\infty} 6^y e^{-6}}{y!}$.

The tedious calculation required to find $P(Y \geq 10)$ can be avoided by using Table 3, Appendix 3, software such as *R*.

From Theorem 3.11,

$$\mu = \lambda = 6, \qquad \sigma^2 = \lambda = 6, \qquad \sigma = \sqrt{6} = 2.45$$

The empirical rule tells us that we should expect *Y* to take values in the interval $\mu \pm 2\sigma$ with a high probability.

Notice that $\mu + 2\sigma = 6 + (2)(2.45) = 10.90$.

The observed number of accidents, $Y = 10$, does not lie more than 2σ from μ, but it is close to the boundary.

Thus, the observed result is not highly improbable, but it may be sufficiently improbable to warrant an investigation. See Exercise 3.210 for the exact probability $P(|Y - \lambda| \leq 2\sigma)$.

Motivations

- The parameters μ and σ are meaningful numerical descriptive measures that locate the center and describe the spread associated with the values of a random variable Y.

- They do not, however, provide a unique characterization of the distribution of Y.

- Many different distributions possess the same means and standard deviations.

- We now consider a set of numerical descriptive measures that (at least under certain conditions) uniquely determine $p(y)$.

Moments of Random Variable

DEFINITION 3.12

The kth *moment of a random variable* Y *taken about the origin* is defined to be $E(Y^k)$ and is denoted by μ'_k.

- $\mu'_1 = E(Y) = \mu$ (First moment about the origin is the mean)
- $\mu'_2 = E(Y^2)$ is employed in Theorem 3.6 for finding σ^2

Central Moment of Random Variable

DEFINITION 3.13

The kth *moment of a random variable* Y *taken about its mean*, or the kth *central moment of* Y, is defined to be $E[(Y - \mu)^k]$ and is denoted by μ_k.

- $\mu_2 = E[(Y - \mu)^2] = \sigma^2$ (second central moment is the variance)
- $\mu_2' = \mathrm{E}(Y^2)$ is employed in Theorem 3.6 for finding σ^2

Central Moment of Random Variable

- Suppose that two random variables Y and Z possess finite moments with

$$\mu'_{1Y} = \mu'_{1Z}, \mu'_{2Y} = \mu'_{2Z}, \ldots, \mu'_{jY} = \mu'_{jZ}$$

  where j can assume any integer value.

- That is, the two random variables possess identical corresponding moments about the origin.
- Under some fairly general conditions, it can be shown that Y and Z have identical probability distributions.
- Thus, a major use of moments is to approximate the probability distribution of a random variable (usually an estimator or a decision maker).

Moment-Generating Functions

DEFINITION 3.14

The *moment-generating function* $m(t)$ *for a random variable* Y is defined to be $m(t) = E(e^{tY})$.We say that a moment-generating function for Y exists if there exists a positive constant b such that $m(t)$ is finite for $|t| \leq b$.

$$m(t) = E(e^{tY}) = \sum_y e^{ty} p(y) = \sum_y \left[1 + ty + \frac{(ty)^2}{2!} + \frac{(ty)^3}{3!} + \cdots\right] p(y) \qquad (\because \text{Taylor's expansion for } e^{ty})$$

$$= \sum_y 1p(y) + t\sum_y yp(y) + \frac{t^2}{2!}\sum_y y^2 p(y) + \frac{t^3}{3!}\sum_y y^3 p(y) + \cdots$$

$$= 1 + t\mu_1' + \frac{t^2}{2!}\mu_2' + \frac{t^3}{3!}\mu_3' + \cdots.$$

- $E(e^{tY})$ is a function of all the moments μ_k' about the origin, for $k = 1, 2, 3, \ldots$.
  - ✓ μ_k' is the coefficient of $t^k/k!$ in the series expansion of $m(t)$

Applications of Moment-Generating Functions

- The moment-generating function possesses two important applications.
  - First, if we can find $E(e^{tY})$, we can find any of the moments for Y.
  - The second (but primary) application of a moment-generating function is to prove that a random variable possesses a particular probability distribution $p(y)$.

Moment-Generating Functions

THEOREM 3.12

If $m(t)$ exists, then for any positive integer k,

$$\left.\frac{d^k m(t)}{dt^k}\right]_{t=0} = m^{(k)}(0) = m'_k.$$

In other words, if you find the kth derivative of $m(t)$ with respect to t and then set $t = 0$, the result will be μ'_k.

Proof:

Moment-Generating Functions

THEOREM 3.12

If $m(t)$ exists, then for any positive integer k,

$$\left.\frac{d^k m(t)}{dt^k}\right]_{t=0} = m^{(k)}(0) = m'_k.$$

In other words, if you find the kth derivative of $m(t)$ with respect to t and then set $t = 0$, the result will be μ'_k.

Proof:

$$m(t) = E(e^{tY}) = \sum_y e^{ty} p(y) = 1 + t\mu'_1 + \frac{t^2}{2!}\mu'_2 + \frac{t^3}{3!}\mu'_3 + \cdots.$$

$$m^{(1)}(t) = \mu'_1 + \frac{2t}{2!}\mu'_2 + \frac{3t^2}{3!}\mu'_3 + \cdots . + \cdots.$$

$$m^{(2)}(t) = \mu'_2 + \frac{2t}{2!}\mu'_3 + \frac{3t^2}{3!}\mu'_4 + \cdots . + \cdots.$$

$$m^{(k)}(t) = \mu'_k + \frac{2t}{2!}\mu'_{k+1} + \frac{3t^2}{3!}\mu'_{k+2} + \cdots . + \cdots.$$

Setting $t = 0$ in each of the above derivatives, we obtain

$$m^{(1)}(0) = \mu'_1, m^{(2)}(0) = \mu'_2, \ldots, m^{(k)}(0) = \mu'_k$$

Moment-Generating Functions : Example

EXAMPLE 3.23

Find the moment-generating function $m(t)$ for a Poisson distributed random variable with mean λ.

Moment-Generating Functions : Example

SOLUTION 3.23

$$
\begin{aligned}
m(t) = E(e^{tY}) &= \sum_{y=0}^{\infty} e^{ty}\, p(y) \\
&= \sum_{y=0}^{\infty} e^{ty} \frac{\lambda^y e^{-\lambda}}{y!} \\
&= e^{-\lambda} \sum_{y=0}^{\infty} \frac{(\lambda e^t)^y}{y!} \\
&= e^{-\lambda} e^{\lambda e^t} \\
&= e^{\lambda(e^t - 1)}
\end{aligned}
$$

Appendix A1.11 to find the Taylor series expansion

$$\sum_{y=0}^{\infty} \frac{(\lambda e^t)^y}{y!} = e^{\lambda e^t}$$

Moment-Generating Functions : Example

EXAMPLE 3.24

Use the moment-generating function of Example 3.23 and Theorem 3.12 to find the mean, μ, and variance, σ^2, for the Poisson random variable.

Moment-Generating Functions : Example

SOLUTION 3.24

$m(t) = E(e^{tY}) = e^{\lambda(e^t-1)}$ when Y has Poisson distribution with the mean rate λ

$$m^{(1)}(t) = \frac{d}{dt}[m(t)] = \frac{d}{dt}\left[e^{\lambda(e^t-1)}\right] = e^{\lambda(e^t-1)} \cdot \lambda e^t$$

$$m^{(2)}(t) = \frac{d^2}{dt^2}[m(t)] = \frac{d^2}{dt^2}\left[e^{\lambda(e^t-1)}\right] = \frac{d}{dt}\left[e^{\lambda(e^t-1)} \cdot \lambda e^t\right] = e^{\lambda(e^t-1)} \cdot (\lambda e^t)^2 + e^{\lambda(e^t-1)} \cdot \lambda e^t$$

$$\mu = m^{(1)}(0) = \left[e^{\lambda(e^t-1)} \cdot \lambda e^t\right]_{t=0} = \lambda$$

$$\mu_2' = m^{(2)}(0) = \left[e^{\lambda(e^t-1)} \cdot (\lambda e^t)^2 + e^{\lambda(e^t-1)} \cdot \lambda e^t\right]_{t=0} = \lambda^2 + \lambda$$

$$\sigma^2 = E(Y^2) - \mu^2 = \mu_2' - \mu^2 = \lambda^2 + \lambda - (\lambda)^2 = \lambda$$

Applications of Moment-Generating Functions

- The moment-generating function possesses two important applications.
  - First, if we can find $E(e^{tY})$, we can find any of the moments for Y.
  - The second (but primary) application of a moment-generating function is to prove that a random variable possesses a particular probability distribution $p(y)$.

Moment-Generating Functions : Example

EXAMPLE 3.25

Suppose that Y is a random variable with moment-generating function $m_Y(t) = e^{3.2(e^t-1)}$
What is the distribution of Y?

Moment-Generating Functions : Example

SOLUTION 3.25

- A Poisson distributed random variable with mean λ is $m(t) = e^{\lambda(e^t - 1)}$
- Note that the moment generating function of Y is exactly equal to the moment-generating function of a Poisson distributed random variable with $\lambda = 3.2$.
- Because moment-generating functions are unique, Y must have a Poisson distribution with mean μ =3.2.

Motivation

- An important class of discrete random variables is one in which Y represents a count and consequently takes integer values: $Y = 0, 1, 2, 3, \ldots$.
    - The binomial, geometric, hypergeometric, and Poisson random variables

- One, involving the theory of queues (waiting lines), is concerned with the number of persons (or objects) awaiting service at a particular point in time.

- Designing manufacturing plants where production consists of a sequence of operations, each taking a different length of time to complete.

- Important in studies of population growth. For example, epidemiologists are interested in the growth of bacterial populations and the growth of the number of persons afflicted by a particular diseases.

Probability-Generating Functions

DEFINITION 3.15

Let Y be an integer-valued random variable for which $P(Y = i) = p_i$, where $i = 0, 1, 2, \ldots$. The *probability-generating function* $P(t)$ for Y is defined to be

$$P(t) = E(t^Y) = p_0 + p_1 t + p_2 t^2 + \cdots = \sum_{i=0}^{\infty} p_i t^i$$

for all values of t such that $P(t)$ is finite.

- In particular, the coefficient of t^i in $P(t)$ is the probability p_i.

Probability-Generating Functions

DEFINITION 3.16

The kth *factorial moment* for a random variable Y is defined to be

$$\mu_{[k]} = E[Y(Y-1)(Y-2)\cdots(Y-k+1)],$$

where k is a positive integer.

- $\mu_{[1]} = E[Y]$
- $\mu_{[2]} = E[Y(Y-1)]$: The second factorial moment is useful in finding the variance for binomial, geometric, and Poisson random variables

Probability-Generating Functions

THEOREM 3.13

If $P(t)$ is the probability-generating function for an integer-valued random

$$\left.\frac{d^k P(t)}{dt^k}\right]_{t=1} = P^{(k)}(1) = m_{[k]}.$$

Proof:

Probability-Generating Functions

THEOREM 3.13

If $P(t)$ is the probability-generating function for an integer-valued random

$$\left.\frac{d^k P(t)}{dt^k}\right]_{t=1} = P^{(k)}(1) = m_{[k]}.$$

Proof:

$$P(t) = p_0 + p_1 t + p_2 t^2 + p_3 t^3 + p_4 t^4 + \cdots,$$

$$P^{(1)}(0) = \frac{dP(1)}{dt} = p_1 + 2p_2 + 3p_3 + 4p_4 + \cdots = \mu_{[1]} = E(Y)$$

$$P^{(2)}(1) = \frac{d^2 P(1)}{d^2 t} = (2)(1)p_2 + (3)(2)p_3 + (4)(3)p_4 + \cdots = \mu_{[2]} = E[Y(Y-1)]$$

$$P^{(k)}(1) = \frac{d^k P(1)}{d^k t} = \sum_{y=k}^{\infty} y(y-1)(y-2)\cdots(y-k+1)\,p(y)$$

$$= E[Y(Y-1)(Y-2)\cdots(Y-k+1)] = \mu_{[k]}$$

Probability-Generating Functions

EXAMPLE 3.26

Find the probability-generating function for a geometric random variable.

Probability-Generating Functions

SOLUTION 3.26

Notice that $p_0 = 0$ because Y cannot assume this value. Then

$$P(t) = E(t^Y) = \sum_{y=1}^{\infty} t^y q^{y-1} p = \sum_{y=1}^{\infty} \left(\frac{p}{q}\right)(qt)^y$$

$$= \frac{p}{q}[qt + (qt)^2 + (qt)^3 + \cdots]$$

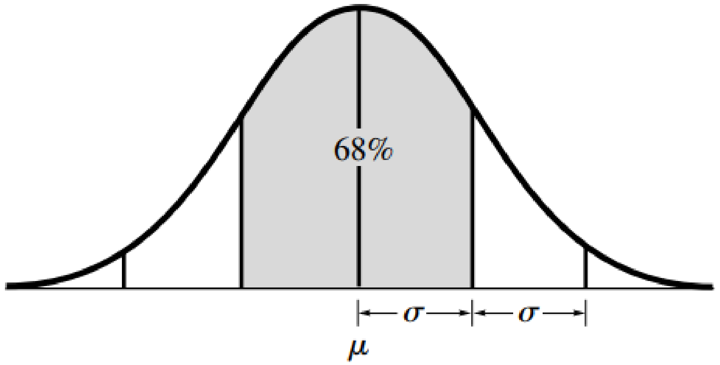
The terms in the series are those of an infinite geometric progression. If $qt < 1$, then

$$P(t) = \frac{p}{q}\left(\frac{qt}{1-qt}\right) = \frac{pt}{1-qt}, \text{ if } t < 1/q$$

(For summation of the series, consult Appendix A1.11.)

Tchebysheff’s Theorem

- If the probability or population histogram is roughly bell-shaped and the mean and variance are known, the empirical rule is of great help in approximating the probabilities of certain intervals.



- However, in many instances, the shapes of probability histograms differ markedly from a mound shape, and the empirical rule may not yield useful approximations to the probabilities of interest.

- Tchebysheff’s theorem, can be used to determine a lower bound for the probability that the random variable Y of interest falls in an interval $\mu \pm k\sigma$.

Tchebysheff's Theorem

THEOREM 3.14

Let Y be a random variable with mean μ and finite variance σ^2. Then, for any constant $k > 0$,

$$P(|Y-\mu| < k\sigma) \geq 1 - \frac{1}{k^2} \quad \text{or} \quad P(|Y-\mu| \geq k\sigma) \leq \frac{1}{k^2}$$

- Tchebysheff's theorem, can be used to determine a lower bound for the probability that the random variable Y of interest falls in an interval $\mu \pm k\sigma$.
- First, the result applies for any probability distribution, whether the probability histogram is bell-shaped or not.
- Second, the results of the theorem are very conservative in the sense that the actual probability that Y is in the interval $\mu \pm k\sigma$ usually exceeds the lower bound for the probability, $1 - \frac{1}{k^2}$, by a considerable amount.

Tchebysheff’s Theorem

EXAMPLE 3.28

The number of customers per day at a sales counter, Y, has been observed for a long period of time and found to have mean 20 and standard deviation 2. The probability distribution of Y is not known. What can be said about the probability that, tomorrow, Y will be greater than 16 but less than 24?

Tchebysheff's Theorem

SOLUTION 3.28

We want to find $P(16 < Y < 24)$. From Theorem 3.14 we know that, for any $k \geq 0$,

$$P(|Y-\mu| < k\sigma) \geq 1 - \frac{1}{k^2}, \text{ or}$$

$$P[(\mu - k\sigma) < Y < (\mu + k\sigma)] \geq 1 - \frac{1}{k^2}$$

Because $\mu = 20$ and $\sigma = 2$, it follows that $\mu - k\sigma = 16$ and $\mu + k\sigma = 24$ if $k = 2$.

Thus,

$$P(16 < Y < 24) = P(\mu - 2\sigma < Y < \mu + 2\sigma) \geq 1 - \frac{1}{(2)^2} = 3/4$$

In other words, tomorrow's customer total will be between 16 and 24 with a fairly high probability (at least $3/4$).

Notice that if σ were 1, k would be 4, and

$$P(16 < Y < 24) = P(\mu - 4\sigma < Y < \mu + 4\sigma) \geq 1 - \frac{1}{(4)^2} = \frac{15}{16}$$

Thus, the value of σ has considerable effect on probabilities associated with intervals.

Continuous Random Variable (Preview)

- Continuous random variable will be discussed in Chapter 4. Below is the preview.

- Suppose a r.v. X can take on a continuum of values each with probability 0.

- Examples:
  - ✓ Pick a number between 0 and 1
  - ✓ Measure the voltage across a heated resistor
  - ✓ Measure the phase of a random sinusoid . . .

- How do we describe probabilities of interesting events?

- Idea:
  - ✓ For discrete r.v., we sum a pmf over points in a set to find its probability.
  - ✓ For continuous r.v., integrate a probability density over a set to find its probability — analogous to mass density in physics (integrate mass density to get the mass)

Continuous Random Variable (Preview)

- A continuous r.v. X can be specified by a probability density function $f_X(x)$ (pdf) such that, for any interval $A \subset \mathbb{R}$,

$$P\{X \in A\} = \int_A f_X(x)\, dx$$

For example, for $A = (a, b]$, the probability can be computed as

$$P\{X \in (a, b]\} = \int_a^b f_X(x)\, dx$$

- Properties of $f_X(x)$:
  - ✓ $f_X(x) \geq 0$
  - ✓ $\int_{-\infty}^{\infty} f_X(x)\, dx = 1$

- Important note: $f_X(x)$ should not be interpreted as the probability that $X = x$, in fact it is not a probability measure, e.g., it can be > 1

Continuous Random Variable (Preview)

- Can relate $f_X(x)$ to a probability using *mean value theorem for integrals*: Fix x and some $\Delta x > 0$. Then provided f_X is continuous over $[x,\ x + \Delta x]$,

$$P\{X \in [x, x + \Delta x]\} = \int_{x}^{x+\Delta x} f_X(\alpha)\, d\alpha$$

$$= f_X(c)\Delta x \text{ for some } x < c < x + \Delta x$$

Now, if Δx is sufficently small, then $P\{X \in [x, x + \Delta x]\} \approx f_X(x)\Delta x$

- Notation: $X \sim f_X(x)$ means that X has pdf $f_X(x)$

Questions