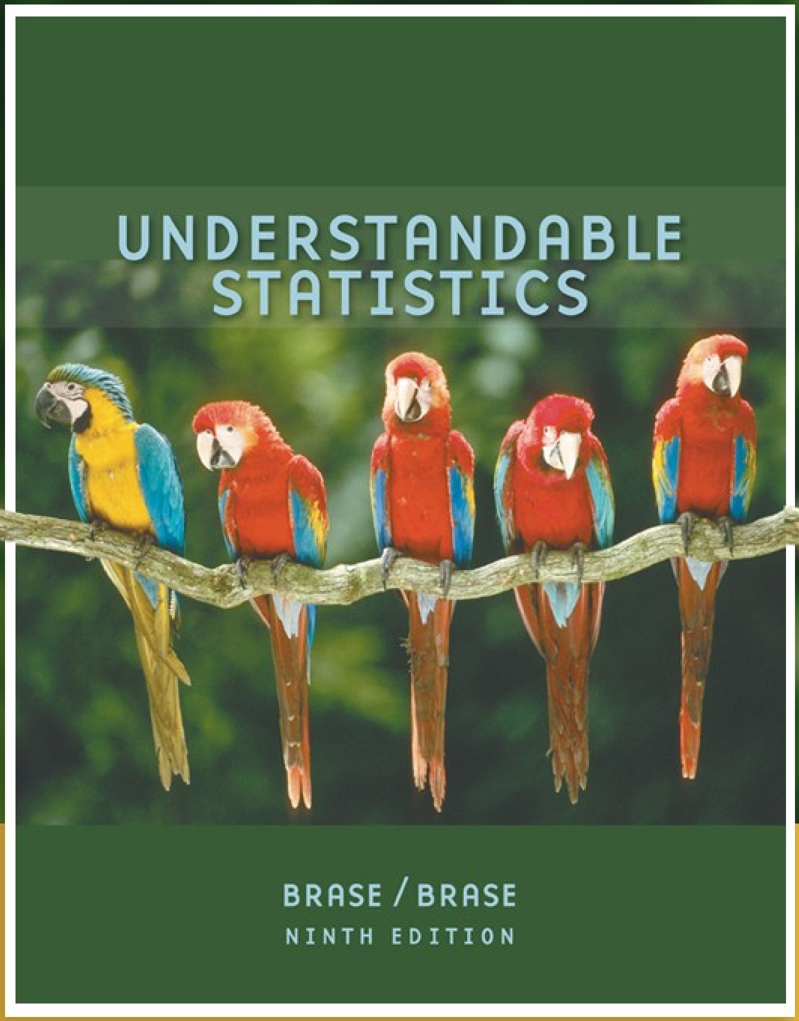


Chapter 5
The Binomial Probability Distribution and Related Topics

Understandable Statistics
Ninth Edition
By Brase and Brase
Prepared by Yixun Shi
Bloomsburg University of Pennsylvania
Edited by José Neville Díaz
University of Puerto Rico at Aguadilla

FOCUS PROBLEM

Personality Preference Types: Introvert or Extrovert?

Isabel Briggs Myers was a pioneer in the study of personality types. Her work has been used successfully in counseling, educational, and industrial settings. In the book *A Guide to the Development and Use of the Myers-Briggs Type Indicators*, by Myers and McCaully, it was reported that based on a very large sample (2282 professors), approximately 45% of all university professors are extroverted.

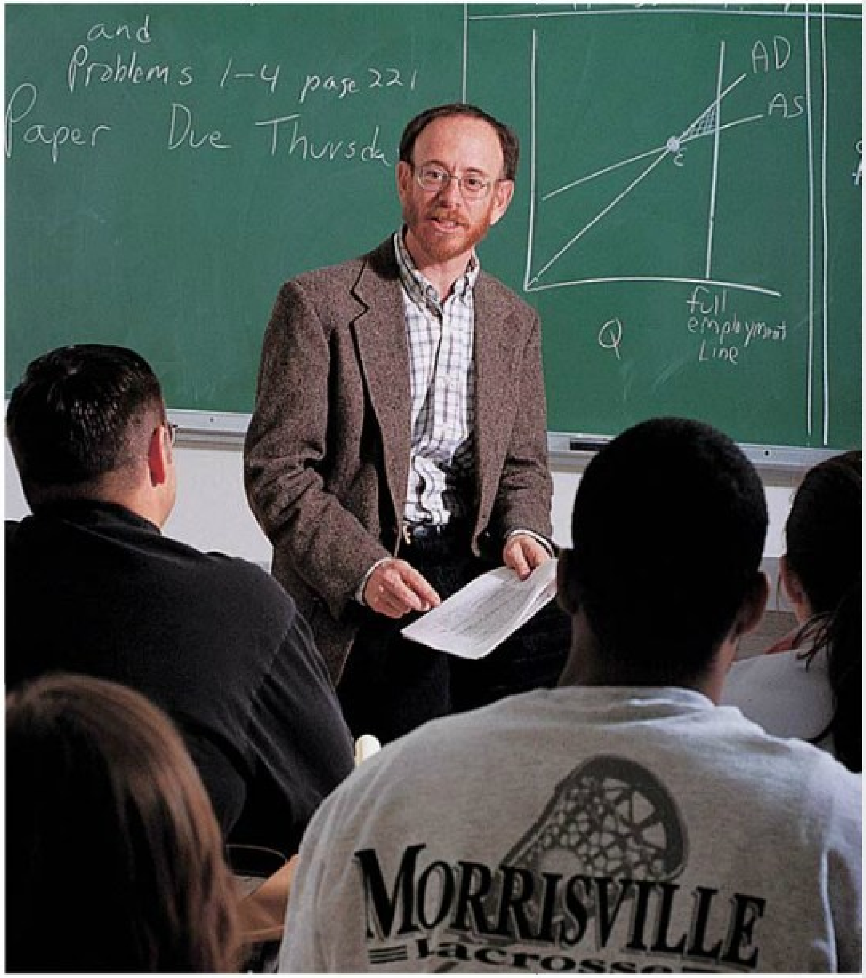
After completing this chapter, you will be able to answer the following questions. Suppose you have classes with six different professors.

(a) What is the probability that all six are extroverts?

(b) What is the probability that none of your professors is an extrovert?

(c) What is the probability that at least two of your professors are extroverts?

(d) In a group of six professors selected at random, what is the *expected number* of extroverts? What is the *standard deviation* of the distribution?



Statistical Experiments and Random Variables

Statistical Experiments – any process by which measurements are obtained.

A quantitative variable, *x*, is a random variable if its value is determined by the outcome of a random experiment.

Random variables can be discrete or continuous

Random Variable

- In probability and statistics, a **random variable** is a variable whose value results from a measurement on some type of random process. Formally, it is a **function.**

- A Random Experiment is an experiment or observation that can be repeated numerous times under the same conditions.
- The outcome of an individual random experiment must be independent and identically distributed. It must in no way be affected by any previous outcome and cannot be predicted with certainty.

Independent and identically distributed (**i.i.d.**) Each random variable has the same probability distribution as the others and all are mutually independent.

Random Variables and Their Probability Distributions

- Discrete random variables – can take on only a countable or finite number of values.
- Continuous random variables – can take on countless values in an interval on the real line
- Probability distributions of random variables – An assignment of probabilities to the specific values or a range of values for a random variable.

- In probability theory and statistics, a probability distribution identifies either the probability of each value of a random variable (when the variable is discrete), or the probability of the value falling within a particular interval (when the variable is continuous).

- The probability distribution describes the range of possible values that a random variable can attain and the probability that the value of the random variable is within any (measurable) subset of that range.

A random variable is a number that is derived from a random process. For example, suppose that we measure the rainfall in July in different cities. One city might have 1.5 inches, another city might have 0.8 inches, and so on. The amount of rainfall is a random variable.

The number of hits Pedro Martinez allows when he pitches in a game
The number of students in a class with birthdays in January
The score that you get when you take the SAT test.

Give some other examples of random variables. Think of some examples that are discrete and some examples that are continuous.

For continuous random variables, we talk about the **probability distribution function**. The probability distribution function pertains to intervals of a continuous random variable.

One example of important distribution function is the uniform distribution function. A random variable is uniformly distributed means that the probability of landing in a particular interval is equal to the size of that interval divided by the size of the entire distribution.

For example, consider a random variable that is uniformly distributed between 0 and 100. The entire distribution has a width of 100. The probability of landing between 0 and 10 is 10/100, or 0.1

Discrete Probability Distributions

- Each value of the random variable has an assigned probability.
- The sum of all the assigned probabilities must equal 1.

Probability Distribution Features

- Since a probability distribution can be thought of as a relative-frequency distribution for a very large n, we can find the mean and the standard deviation.

- When viewing the distribution in terms of the population, use μ for the mean and σ for the standard deviation.

EXAMPLE 1 Discrete probability distribution

Dr. Mendoza developed a test to measure boredom tolerance. He administered it to a group of 20,000 adults between the ages of 25 and 35. The possible scores were 0, 1, 2, 3, 4, 5, and 6, with 6 indicating the highest tolerance for boredom. The test results for this group are shown in Table 5-1.

(a) If a subject is chosen at random from this group, the probability that he or she will have a score of 3 is 6000/20,000, or 0.30. In a similar way, we can use relative frequencies to compute the probabilities for the other scores (Table 5-2). These probability assignments make up the probability distribution. Notice that the scores are mutually exclusive: No one subject has two scores. The sum of the probabilities of all the scores is 1.

TABLE 5-1 **Boredom Tolerance Test Scores for 20,000 Subjects**

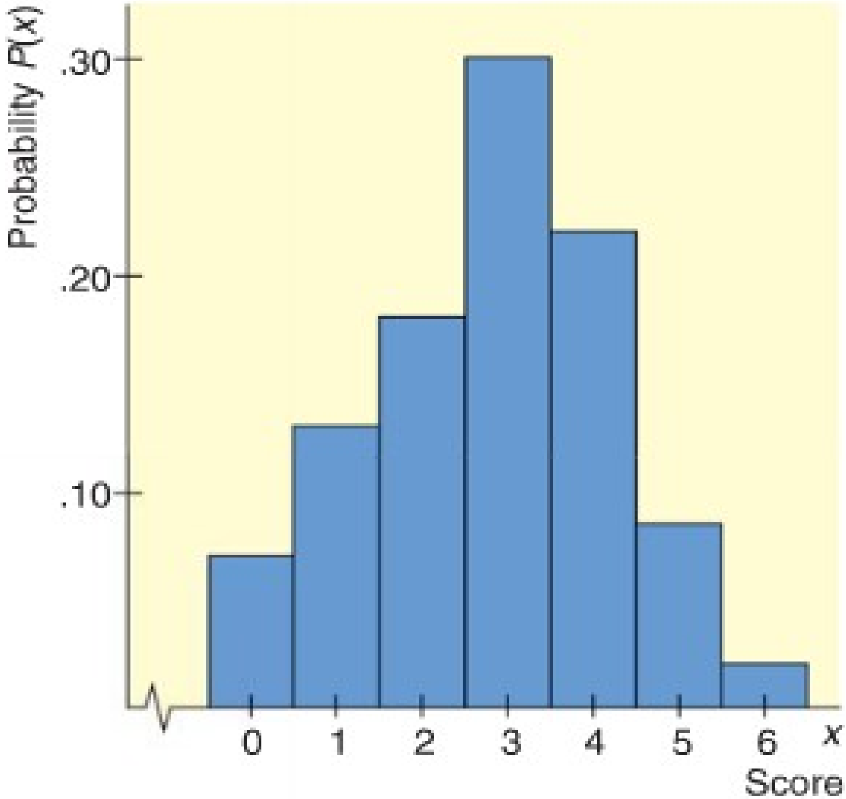
Score	Number of Subjects
0	1400
1	2600
2	3600
3	6000
4	4400
5	1600
6	400

TABLE 5-2 **Probability Distribution of Scores on Boredom Tolerance Test**

Score x	Probability $P(x)$
0	0.07
1	0.13
2	0.18
3	0.30
4	0.22
5	0.08
6	0.02
	$\Sigma P(x) = 1$

FIGURE 5-1

Graph of the Probability Distribution of Test Scores



P(5 o 6) = ?

GUIDED EXERCISE 2 Discrete probability distribution

One of the elementary tools of cryptanalysis (the science of code breaking) is to use relative frequencies of occurrence of different letters in the alphabet to break standard English alphabet codes. Large samples of plain text such as newspaper stories generally yield about the same relative frequencies for letters. A sample 1000 letters long yielded the information in Table 5-3.

(a) Use the relative frequencies to compute the omitted probabilities in Table 5-3.

→ Table 5-4 shows the completion of Table 5-3.

TABLE 5-3 Frequencies of Letters in a 1000-Letter Sample

Letter	Freq.	Prob.	Letter	Freq.	Prob.
A	73	———	N	78	0.078
B	9	0.009	O	74	———
C	30	0.030	P	27	0.027
D	44	0.044	Q	3	0.003
E	130	———	R	77	0.077
F	28	0.028	S	63	0.063
G	16	0.016	T	93	0.093
H	35	0.035	U	27	———
I	74	———	V	13	0.013
J	2	0.002	W	16	0.016
K	3	0.003	X	5	0.005
L	35	0.035	Y	19	0.019
M	25	0.025	Z	1	0.001

Source: Elementary Cryptanalysis: A Mathematical Approach, by Abraham Sinkov. Copyright © 1968 by Yale University. Reprinted by permission of Random House, Inc.

TABLE 5-4 Entries for Table 5-3

Letter	Relative Frequency	Probability
A	$\frac{73}{1,000}$	0.073
E	$\frac{130}{1,000}$	0.130
I	$\frac{74}{1,000}$	0.074
O	$\frac{74}{1,000}$	0.074
U	$\frac{27}{1,000}$	0.027

(b) Do the probabilities of all the individual letters add up to 1?

(c) If a letter is selected at random from a newspaper story, what is the probability that the letter will be a vowel?

Means and Standard Deviations for Discrete Probability Distributions

The **mean** and the **standard deviation of a discrete population probability distribution** are found by using these formulas:

$\mu = \Sigma x P(x)$; μ is called the **expected value** of x

$\sigma = \sqrt{\Sigma (x - \mu)^2 P(x)}$ is called the standard deviation of x

where x is the value of a random variable,
$P(x)$ is the probability of that variable, and
the sum Σ is taken for all the values of the random variable.

Note: μ is the *population mean* and σ is the underlying *population standard deviation* because the sum Σ is taken over *all* values of the random variable (i.e., the entire sample space).

Expected value, standard deviation

Are we influenced to buy a product by an ad we saw on TV? National Infomercial Marketing Association determined the number of times *buyers* of a product watched a TV infomercial *before* purchasing the product. The results are shown here:

Number of Times Buyers Saw Infomercial	1	2	3	4	5*
Percentage of Buyers	27%	31%	18%	9%	15%

*This category was 5 or more, but will be treated as 5 in this example.

We can treat the information shown as an estimate of the probability distribution because the events are mutually exclusive and the sum of the percentages is 100%. Compute the mean and standard deviation of the distribution.

Linear Functions of Random Variables

Let *a* and *b* be constants.

Let *x* be a random variable.

$L = a + bx$ is a linear function of *x*.

Finding μ and σ for Linear Functions of x

Let x be a random variable with mean μ and standard deviation σ. Then the **linear function** $L = a + bx$ has mean, variance, and standard deviation as follows:

$$\mu_L = a + b\mu$$

$$\sigma_L^2 = b^2\sigma^2$$

$$\sigma_L = \sqrt{b^2\sigma^2} = |b|\sigma$$

Independent Random Variables

Let *x1* and *x2* be random variables.
Then the random variables are independent
if any event of *x1* is independent of any event of *x2*.

Combining Random Variables

Let x_1 and x_2 be independent random variables with respective means μ_1 and μ_2, and variances σ_1^2 and σ_2^2. For the **linear combination** $W = ax_1 + bx_2$, the mean, variance, and standard deviation are as follows:

$$\mu_W = a\mu_1 + b\mu_2$$
$$\sigma_W^2 = a^2\sigma_1^2 + b^2\sigma_2^2$$
$$\sigma_W = \sqrt{a^2\sigma_1^2 + b^2\sigma_2^2}$$

Linear combinations of independent random variables

Let x_1 and x_2 be independent random variables with respective means $\mu_1 = 75$ and $\mu_2 = 50$, and standard deviations $\sigma_1 = 16$ and $\sigma_2 = 9$.

(a) Let $L = 3 + 2x_1$. Compute the mean, variance, and standard deviation of L.

(b) Let $W = x_1 + x_2$. Find the mean, variance, and standard deviation of W.

(c) Let $W = x_1 - x_2$. Find the mean, variance, and standard deviation of W.

(d) Let $W = 3x_1 - 2x_2$. Find the mean, variance, and standard deviation of W.

Binomial Experiments

There are a fixed number of trials.

This is denoted by n.

The n trials are independent and repeated under identical conditions.

Each trial has two outcomes:

S = *success* F = *failure*

Binomial Experiments

- For each trial, the probability of success, p, remains the same. Thus, the probability of failure is $1 - p = q$.

- The central problem is to determine *the probability of* r *successes out of* n *trials.*

Determining Binomial Probabilities

- Use the Binomial Probability Formula.
- Use Table 3 of Appendix II.
- Use technology.

Binomial Probability Formula

Formula for the binomial probability distribution

$$P(r) = \frac{n!}{r!(n-r)!} p^r q^{n-r} = C_{n,r} p^r q^{n-r}$$

where n = number of trials

p = probability of success on each trial

$q = 1 - p$ = probability of failure on each trial

r = random variable representing the number of successes out of n trials $(0 \le r \le n)$

$!$ = factorial notation. Recall from Section 4.3 that the factorial symbol $n!$ designates the product of all the integers between 1 and n. For instance, $4! = 4 \cdot 3 \cdot 2 \cdot 1 = 24$. Special cases are $1! = 1$ and $0! = 0$.

$C_{n,r} = \frac{n!}{r!(n-r)!}$ is the binomial coefficient. Table 2 of Appendix II gives values of $C_{n,r}$ for select n and r. Many calculators have a key designated nCr that gives the value of $C_{n,r}$ directly.

Using the Binomial Table

Locate the number of trials, n.

Locate the number of successes, r.

Follow that row to the right to the corresponding p column.

		p															
n	*r*	.01	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50	.55	.60	.65	.70	.75
8	0	.923	.663	.430	.272	.168	.100	.058	.032	.017	.008	.004	.002	.001	.000	.000	.000
	1	.075	.279	.383	.385	.336	.267	.198	.137	.090	.055	.031	.016	.008	.003	.001	.000
	2	.003	.051	.149	.238	.294	.311	.296	.259	.209	.157	.109	.070	.041	.022	.010	.004
	3	.000	.005	.033	.084	.147	.208	.254	.279	.279	.257	.219	.172	.124	.081	.047	.023
	4	.000	.000	.005	.018	.046	.087	.136	.188	.232	.263	.273	.263	.232	.188	.136	.087
	5	.000	.000	.000	.003	.009	.023	.047	.081	.124	.172	.219	.257	.279	.279	.254	.208
	6	.000	.000	.000	.000	.001	.004	.010	.022	.041	.070	.109	.157	.209	.259	.296	.311
	7	.000	.000	.000	.000	.000	.000	.001	.003	.008	.016	.031	.055	.090	.137	.198	.267
	8	.000	.000	.000	.000	.000	.000	.000	.000	.001	.002	.004	.008	.017	.032	.058	.100
9	0	.914	.630	.387	.232	.134	.075	.040	.021	.010	.005	.002	.001	.000	.000	.000	.000
	1	.083	.299	.387	.368	.302	.225	.156	.100	.060	.034	.018	.008	.004	.001	.000	.000
	2	.003	.063	.172	.260	.302	.300	.267	.216	.161	.111	.070	.041	.021	.010	.004	.001
	3	.000	.008	.045	.107	.176	.234	.267	.272	.251	.212	.164	.116	.074	.042	.021	.009
	4	.000	.001	.007	.028	.066	.117	.172	.219	.251	.260	.246	.213	.167	.118	.074	.039
	5	.000	.000	.001	.005	.017	.039	.074	.118	.167	.213	.246	.260	.251	.219	.172	.117
	6	.000	.000	.000	.001	.003	.009	.021	.042	.074	.116	.164	.212	.251	.272	.267	.234
	7	.000	.000	.000	.000	.000	.001	.004	.010	.021	.041	.070	.111	.161	.216	.267	.300

Recall for the sharpshooter example, *n* = 8, *r* = 6, *p* = 0.7
So the probability she hits exactly 6 targets is 0.296, as expected

Binomial Probabilities

- At times, we will need to calculate other probabilities:
  - $P(r < k)$
  - $P(r \leq k)$
  - $P(r > k)$
  - $P(r \geq k)$

Where k is a specified value less than or equal to the number of trials, n.

Graphing a Binomial Distribution

PROCEDURE

How to graph a binomial distribution

1. Place r values on the horizontal axis.
2. Place $P(r)$ values on the vertical axis.
3. Construct a bar over each r value extending from $r - 0.5$ to $r + 0.5$. The height of the corresponding bar is $P(r)$.

Mean and Standard Deviation of a Binomial Distribution

$$\mu = np$$

$$\sigma = \sqrt{npq}$$

Critical Thinking

- Unusual values – For a binomial distribution, it is unusual for the number of successes *r* to be more than 2.5 standard deviations from the mean.
  - This can be used as an indicator to determine whether a specified number of *r* out of *n* trials in a binomial experiment is unusual.

Quota Problems

We can use the binomial distribution table "backwards" to solve for a minimum number of trials.

In these cases, we know *r* and *p*

We use the table to find an *n* that satisfies our required probability

The Geometric Distribution

- Suppose that rather than repeat a fixed number of trials, we repeat the experiment until the first *success*.
- Examples:
  - Flip a coin until we observe the first head
  - Roll a die until we observe the first 5
  - Randomly select DVDs off a production line until we find the first defective disk

Geometric probability distribution

$$P(n) = p(1 - p)^{n - 1}$$

where n is the number of the trial on which the *first success* occurs ($n = 1, 2, 3, \ldots$) and p is the probability of success on each trial. *Note:* p must be the same for each trial.

Using some mathematics (involving infinite series), it can be shown that the **population mean** and **standard deviation** of the geometric distribution are

$$\mu = \frac{1}{p} \quad \text{and} \quad \sigma = \frac{\sqrt{1 - p}}{p}$$

A rat has to choose between 5 doors, one of which contains chocolate.If the rat chooses the wrong door, it is returned to the starting point and chooses again, and continues until it gets the chocolate. Let X be the serial number of the trial on which the chocolate is found.

(a) Find the probability function of X

(b) What is the expectation of X?

Solution

The Poisson Distribution

This distribution is used to model the number of “rare” events that occur in a time interval, volume, area, length, etc…

Examples:

Number of auto accidents during a month

Number of diseased trees in an acre

Number of customers arriving at a bank

The Poisson Distribution

Poisson distribution

Let λ (Greek letter lambda) be the mean number of successes over time, volume, area, and so forth. Let r be the number of successes ($r = 0, 1, 2, 3, \ldots$) in a corresponding interval of time, volume, area, and so forth. Then the probability of r successes in the interval is

$$P(r) = \frac{e^{-\lambda}\lambda^r}{r!}$$

where e is approximately equal to 2.7183.

Using some mathematics (involving infinite series), it can be shown that the **population mean** and **standard deviation** of the Poisson distribution are

$$\mu = \lambda \qquad \text{and} \qquad \sigma = \sqrt{\lambda}$$

Finding Poisson Probabilities Using the Table

We can use Table 4 of Appendix II instead of the formula.

1) Find λ at the top of the table.
2) Find r along the left margin of the table.

Using the Poisson Table

	λ									
r	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9	4.0
0	.0450	.0408	.0369	.0334	.0302	.0273	.0247	.0224	.0202	.0183
1	.1397	.1304	.1217	.1135	.1057	.0984	.0915	.0850	.0789	.0733
2	.2165	.2087	.2008	.1929	.1850	.1771	.1692	.1615	.1539	.1465
3	.2237	.2226	.2209	.2186	.2158	.2125	.2087	.2046	.2001	.1954
4	.1734	.1781	.1823	.1858	.1888	.1912	.1931	.1944	.1951	.1954
5	.1075	.1140	.1203	.1264	.1322	.1377	.1429	.1477	.1522	.1563
6	.0555	.0608	.0662	.0716	.0771	.0826	.0881	.0936	.0989	.1042
7	.0246	.2078	.0312	.0348	.0385	.0425	.0466	.0508	.0551	.0595
8	.0095	.0111	.0129	.0148	.0169	.0191	.0215	.0241	.0269	.0298
9	.0033	.0040	.0047	.0056	.0066	.0076	.0089	.0102	.0116	.0132
10	.0010	.0013	.0016	.0019	.0023	.0028	.0033	.0039	.0045	.0053
11	.0003	.0004	.0005	.0006	.0007	.0009	.0011	.0013	.0016	.0019
12	.0001	.0001	.0001	.0002	.0002	.0003	.0003	.0004	.0005	.0006
13	.0000	.0000	.0000	.0000	.0001	.0001	.0001	.0001	.0002	.0002
14	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001

Recall, λ = 4

***r* = 0**

***r* = 4**

***r* = 7**

Poisson Approximation to the Binomial

PROCEDURE

How to approximate binomial probabilities using Poisson probabilities

Suppose you have a binomial distribution with

n = number of trials
r = number of successes
p = probability of success on each trial

If $n \geq 100$ and $np < 10$, then r has a binomial distribution that is approximated by a Poisson distribution with $\lambda = np$.

$$P(r) \approx \frac{e^{-\lambda}\lambda^r}{r!}$$

Note: $\lambda = np$ is the expected value of the binomial distribution.

Suppose the average number of lions seen on a 1-day safari is 5. What is the probability that tourists will see fewer than four lions on the next 1-day safari?

Solution: This is a Poisson experiment in which we know the following:

Vehicles pass through a junction on a busy road at an average rate of 300 per hour.

Find the probability that none passes in a given minute.

What is the expected number passing in two minutes?

Find the probability that this expected number actually pass through in a given two-minute period.

Solution:

Lamda =300/1hr   Nuevo lamda=300/60=5 1min

a) P(0)=.0067

b) 5x2=10

c) P(10)=.125

Binomial Approximation by Poisson

A manufacturer of Christmas tree bulbs knows that 2% of its bulbs are defective. Approximate the probability that a box of 100 of these bulbs contains at most three defective bulbs. Assuming independence, we have binomial distribution with parameters p=0.02 and n=100.

The Poisson distribution with gives using the binomial distribution, we obtain, after some tedious calculations,   Hence, in this case, the Poisson approximation is extremely close to the true value, but much easier to find.

A company makes electric motors. The probability an electric motor is defective is 0.01. What is the probability that a sample of 300 electric motors will contain exactly 5 defective motors?