Marmara University, 2021

Probability and Statistics

Subject 5

Several Useful Discrete Distributions

Mujdat Soyturk, Ph.D.
Associate Professor

Contents

- The Binomial Probability Distribution
- The Poisson Probability Distribution
- The Hypergeometric Probability Distribution

Most parts of the slides are derived from the textbook: "Mendenhall, Beaver, Beaver, *Introduction to Probability and Statistics*, 14th Ed., Brooks/Cole, Cengage Learning, 2013"

Introduction

- Discrete random variables take on only a finite or countably infinite number of values.
- Three discrete probability distributions serve as models for a large number of practical applications:

✓The **binomial** random variable

✓The **Poisson** random variable

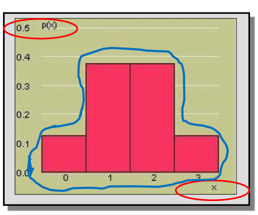
✓The **hypergeometric** random variable

Introduction

- These three discrete random variables are often used to describe the number of occurrences of a specified event in a *fixed number of trials* or *a fixed unit of time or space*.

The Binomial Random Variable

- The **coin-tossing experiment** is a simple example of a **binomial random variable.**
- Toss a fair coin $n = 3$ times and record x = number of heads.



x	$p(x)$
0	1/8
1	3/8
2	3/8
3	1/8

The Binomial Random Variable

Many situations in real life resemble the coin toss, but the coin is not necessarily fair, so that $P(H) \neq 1/2$.

Example: A geneticist samples 10 people and counts the number who have a gene linked to Alzheimer's disease.

- Coin: Person
- Head: Has gene
- Tail: Doesn't have gene
- Number of tosses: $n = 10$
- P(H): P(has gene) = proportion in the population who have the gene.

The Binomial Experiment

1. The experiment consists of n **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 p and **remains constant** from trial to trial. The probability of failure is $q = 1 - p$.
4. The trials are **independent.**
5. We are interested in x**, the number of successes in** n **trials.**

Binomial or Not?

Very few real life applications satisfy these requirements exactly.

- Select two people from the U.S. population, and suppose that 15% of the population has the Alzheimer's gene.
  - For the first person, p = P(gene) = .15
  - For the second person, $p \approx$ P(gene) = .15, even though one person has been removed from the population.

Binomial or Not?

- Keep in mind and check for the below:
  - When the sample (the n identical trials) came from a large population, the probability of success p stayed about the same from trial to trial. **Binomial**
  - When the population size N was small, the probability of success p changed quite dramatically from trial to trial, and the experiment *was not binomial*.
- Rule of Thumb:
  - If $n/N \geq .05$, the experiment *is not binomial*.

The Binomial Probability Distribution

- For a binomial experiment with n trials and probability p of success on a given trial, the probability of k successes in n trials is

$$P(x=k) = C_k^n p^k q^{n-k} = \frac{n!}{k!(n-k)!} p^k q^{n-k} \text{ for } k = 0,1,2,\ldots n.$$

$$\text{Recall} \quad C_k^n = \frac{n!}{k!(n-k)!}$$

$$\text{with } n! = n(n-1)(n-2)\ldots(2)1 \text{ and } 0! \equiv 1.$$

The Mean and Standard Deviation

- For a binomial experiment with n trials and probability p of success on a given trial, the measures of center and spread are:

$$\text{Mean}: \mu = np$$

$$\text{Variance}: \sigma^2 = npq$$

$$\text{Standard deviation}: \sigma = \sqrt{npq}$$

Ex. 5.3 pp.179

Example

A marksman hits a target 80% of the time. He fires five shots at the target. What is the probability that exactly 3 shots hit the target?

***n* = 5** ***success* = hit** ***p* = .8** ***x* = # of hits**

$$P(x=3) = C_3^n p^3 q^{n-3} = \frac{5!}{3!2!}(.8)^3(.2)^{5-3}$$

$$= 10(.8)^3(.2)^2 = .2048$$

Example

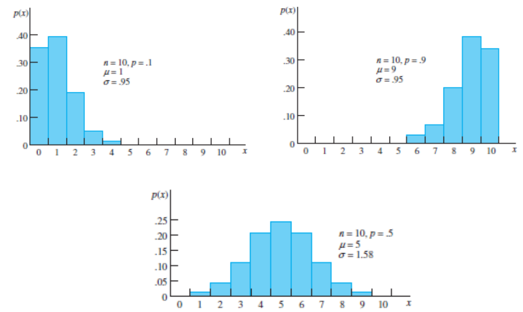
What is the probability that more than 3 shots hit the target?

$$P(x>3)=C_4^5 p^4 q^{5-4}+C_5^5 p^5 q^{5-5}$$

$$=\frac{5!}{4!1!}(.8)^4(.2)^1+\frac{5!}{5!0!}(.8)^5(.2)^0$$

$$=5(.8)^4(.2)+(.8)^5=.7373$$

Example



Cumulative Probability Tables

You can use the **cumulative probability tables** to find probabilities for selected binomial distributions.

1) Find the table for the correct value of ***n***.

2) Find the column for the correct value of ***p***.

3) The row marked "*k*" gives the cumulative probability, $P(x \le k) = P(x = 0) + \ldots + P(x = k)$

Example

k	*p* = .80
0	.000
1	.007
2	.058
3	.263
4	.672
5	1.000

What is the probability that *exactly 3 shots* hit the target?

P(*x* = 3) = P(*x* ≤ 3) – P(*x* = 2) – P(*x* = 1) – P(*x* = 0)
P(*x* = 3) = P(*x* ≤ 3) – P(*x* ≤ 2)
= .263 - .058 = .205

Check from formula: P(*x* = 3) = .2048

n = 5

	p													
k	.01	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	.99	*k*
0	.951	.774	.590	.328	.168	.078	.031	.010	.002	.000	.000	.000	.000	0
1	.999	.977	.919	.737	.528	.337	.188	.087	.031	.007	.000	.000	.000	1
2	1.000	.999	.991	.942	.837	.683	.500	.317	.163	.058	.009	.001	.000	2
3	1.000	1.000	1.000	.993	.969	.913	.812	.663	.472	.263	.081	.023	.001	3
4	1.000	1.000	1.000	1.000	.998	.990	.969	.922	.832	.672	.410	.226	.049	4
5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	5

Example

k	*p* = .80
0	.000
1	.007
2	.058
3	.263
4	.672
5	1.000

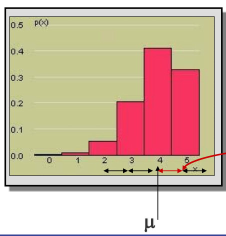
What is the probability that more than 3 shots hit the target?

P(*x* > 3) = 1 - P(*x* ≤ 3)
= 1 - .263 = .737

Check from formula: P(*x* > 3) = .7373

Example

Here is the probability distribution for *x* = **number of hits**. What are the mean and standard deviation for *x*?



$$\text{Mean}: \mu = np = 5(.8) = 4$$

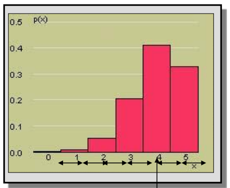
$$\text{Standard deviation}: \sigma = \sqrt{npq}$$

$$= \sqrt{5(.8)(.2)} = .89$$

Example

Would it be unusual to find that none of the shots hit the target?

$\mu = 4; \sigma = .89$



μ

The value $x = 0$ lies

$$z = \frac{x-\mu}{\sigma} = \frac{0-4}{.89} = -4.49$$

more than 4 standard deviations below the mean. **Very unusual.**

The Poisson Random Variable

- The Poisson random variable x is a model for data that represent the *number of occurrences* of a specified event in a *given unit of time or space.*

- **Examples:**
  - The number of calls received by a switchboard during a given period of time.
  - The number of machine breakdowns in a day
  - The number of traffic accidents at a given intersection during a given time period.

The Poisson Random Variable

- The only assumption needed to model experiments is that the counts or events occur randomly and independently of one another.

The Poisson Probability Distribution

- x is the number of events that occur in a period of time or space during which an average of μ such events can be expected to occur. The probability of k occurrences of this event is

$$P(x = k) = \frac{\mu^k e^{-\mu}}{k!}$$

For values of k = 0, 1, 2, … The mean and standard deviation of the Poisson random variable are

Mean: μ

Standard deviation: $\sigma = \sqrt{\mu}$

where e = 2.71828…

Example

The average number of traffic accidents on a certain section of highway is two per week. Find the probability of exactly one accident during a one-week period. **Mean, μ = 2 accidents/week**

$$P(x=1) = \frac{\mu^k e^{-\mu}}{k!} = \frac{2^1 e^{-2}}{1!} = 2e^{-2} = .2707$$

Cumulative Probability Tables

You can use the **cumulative probability tables** to find probabilities for selected Poisson distributions.

1) Find the column for the correct value of μ.
2) The row marked "k" gives the cumulative probability, $P(x \le k) = P(x = 0) + \ldots + P(x = k)$

Example

k	$\mu = 2$
0	.135
1	.406
2	.677
3	.857
4	.947
5	.983
6	.995
7	.999
8	1.000

What is the probability that there is exactly 1 accident?

k	2.0	2.5	3.0	3.5	4.0	4.5
0	.135	.082	.055	.033	.018	.011
1	.406					
2	.677					
3	.857					
4	.947					
5	.983					
6	.995					
7	.999					
8	1.000	.999	.996	.990	.979	.960
9		1.000	.999	.997	.992	.983
10			1.000			
11						
12						
13						

P(x = 1) = P($x \le 1$) – P($x \le 0$)
= .406 - .135
= .271

Check from formula:
P(x = 1) = .2707

Example

k	$\mu = 2$
0	.135
1	.406
2	.677
3	.857
4	.947
5	.983
6	.995
7	.999
8	1.000

What is the probability that 8 or more accidents happen?

P($x \ge 8$) = 1 - P(x < 8)
= 1 – P($x \le 7$)
= 1 - .999 = .001

This would be very unusual (small probability) since x = 8 lies

$$z = \frac{x-\mu}{\sqrt{\mu}} = \frac{8-2}{1.414} = 4.24$$

standard deviations above the mean.

The Poisson Approximation to the Binomial Distribution

- When the number of trial for the Binomial experiment is high, then it is not easy to compute the probabilities and find out the tables for the cumulative probabilities.
- We *can estimate* binomial probabilities with the Poisson when *n* is large and *p* is small.
- The Poisson probability distribution provides a simple, easy-to-compute, and accurate approximation to binomial probabilities when *n is large* and *μ = np is small*, preferably with *np < 7*.

The Poisson Approximation to the Binomial Distribution

Example: Suppose a life insurance company insures the lives of 5000 men aged 42. If actuarial studies show the probability that any 42-year-old man will die in a given year to be .001, find the exact probability that the company will have to pay x = 4 claims during a given year.

1st Solution: $P(x=4) = p(4) = \frac{5000!}{4!4996!}(.001)^4(.999)^{4996}$

2nd Solution (use Poisson Distribution):

μ = n.p = 5000 x 0.001 = 5

$$p(4) \approx \frac{\mu^4 e^{-\mu}}{4!} = \frac{5^4 e^{-5}}{4!} = \frac{(625)(.006738)}{24} = .175$$

or from Table (Poisson Cumulative Distribution) with μ = 5:

$$p(4) = P(x \le 4) - P(x \le 3) = .440 - .265 = .175$$

The Hypergeometric Probability Distribution

- Remember the requirements in a binomial experiment.
- If the number of elements in the population is small in relation to the sample size ($n/N \ge 0.05$), the probability of a success for a given trial is dependent on the outcomes of preceding trials.
- Then the number of x of successes follows what is known as a hypergeometric probability distribution.

The Hypergeometric Probability Distribution

- The "candies problems" from Chapter 4 are modeled by the hypergeometric distribution.
- A bowl contains M red candies and N-M blue candies. Select n candies from the bowl and record x the number of red candies selected. Define a "red candies" to be a "success".

The probability of exactly k successes in n trials is

$$P(x=k) = \frac{C_k^M C_{n-k}^{N-M}}{C_n^N}$$

M → successes, ***N-M*** → failures, ***n*** → size of the random sample space

The Mean and Variance

The mean and variance of the hypergeometric random variable x resemble the mean and variance of the binomial random variable:

$$\text{Mean}: \mu = n\left(\frac{M}{N}\right)$$

$$\text{Variance}: \sigma^2 = n\left(\frac{M}{N}\right)\left(\frac{N-M}{N}\right)\left(\frac{N-n}{N-1}\right)$$

Example

A package of 8 AA batteries contains 2 batteries that are defective. A student randomly selects four batteries and replaces the batteries in his calculator. What is the probability that all four batteries work?

Success = working battery

N = 8

M = 6

***n* = 4**

$$P(x=4) = \frac{C_4^6 C_0^2}{C_4^8}$$

$$= \frac{6(5)/2(1)}{8(7)(6)(5)/4(3)(2)(1)} = \frac{15}{70}$$

Example

What are the mean and variance for the number of batteries that work?

$$\mu = n\left(\frac{M}{N}\right) = 4\left(\frac{6}{8}\right) = 3$$

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

$$= 4\left(\frac{6}{8}\right)\left(\frac{2}{8}\right)\left(\frac{4}{7}\right) = .4286$$

Key Concepts

I. The Binomial Random Variable

1. Five characteristics: *n* identical independent trials, each resulting in either success ***S*** or failure ***F***; probability of success is ***p*** and remains constant from trial to trial; and ***x*** is the number of successes in ***n*** trials.
2. Calculating binomial probabilities
   a. Formula: $P(x=k) = C_k^n p^k q^{n-k}$
   b. Cumulative binomial tables
   c. Individual and cumulative probabilities using Minitab
3. Mean of the binomial random variable: $\mu = np$
4. Variance and standard deviation: $\sigma^2 = npq$ and $\sigma = \sqrt{npq}$

Key Concepts

II. The Poisson Random Variable

1. The number of events that occur in a period of time or space, during which an average of μ such events are expected to occur
2. Calculating Poisson probabilities
   a. Formula: $P(x=k) = \dfrac{\mu^k e^{-\mu}}{k!}$
   b. Cumulative Poisson tables
   c. Individual and cumulative probabilities using Minitab
3. Mean of the Poisson random variable: $E(x) = \mu$
4. Variance and standard deviation: $\sigma^2 = \mu$ and $\sigma = \sqrt{\mu}$
5. Binomial probabilities can be approximated with Poisson probabilities when $np < 7$, using $\mu = np$.

Key Concepts

III. The Hypergeometric Random Variable

1. The number of successes in a sample of size *n* from a finite population containing *M* successes and *N* - *M* failures
2. Formula for the probability of *k* successes in *n* trials:
   $$P(x=k) = \frac{C_k^M C_{n-k}^{N-M}}{C_n^N}$$
3. Mean of the hypergeometric random variable: $\mu = n\left(\frac{M}{N}\right)$
4. Variance and standard deviation:
   $$\sigma^2 = n\left(\frac{M}{N}\right)\left(\frac{N-M}{N}\right)\left(\frac{N-n}{N-1}\right)$$