Ch5. Discrete Probability Distributions

The variables that we discussed in chapter1 and 2 are now redefined as random variables, whose values depend on a chance or random event. Probability is used as a tool to create probability distributions, which serve as models for discrete random variables. Discrete random variables are discussed in general in this chapter. In addition, there are three important discrete random variables- the binomial, the Poisson, and the hypergeometric - that serve as models for many practical applications. These random variables are, often used to describe the number of occurrences of an event in a fixed number of trials or a fixed unit of time or space, are discussed in detail.

Random Variables

- A quantitative variable x is a **random variable** if the value that it assumes, corresponding to the outcome of an experiment is a chance or random event.
- Random variables can be **discrete** or **continuous.**

Example

- ✓ x = Number of defects on a randomly selected piece of furniture
- ✓ x = SAT score for a randomly selected college applicant
- ✓ x = number on the upper face of a randomly tossed die

Probability Distributions for Discrete Random Variables

- The **probability distribution for a discrete random variable x** resembles the relative frequency distributions we constructed in Chapter 1.
- It is a graph, table or formula that gives the possible values of x and the probability $p(x)$ associated with each value.
- Requirements :

$$0 \leq p(x) \leq 1 \text{ and } \sum p(x) = 1$$

Example

- Toss two fair coins and let x equal the number of heads observed. Find the probability distribution for x.

Sol)

Simple Event	Coin 1	Coin 2	$P(E_i)$	x
E_1	H	H	1/4	2
E_2	H	T	1/4	1
E_3	T	H	1/4	1
E_4	T	T	1/4	0

$P(x = 0) = 1/4$

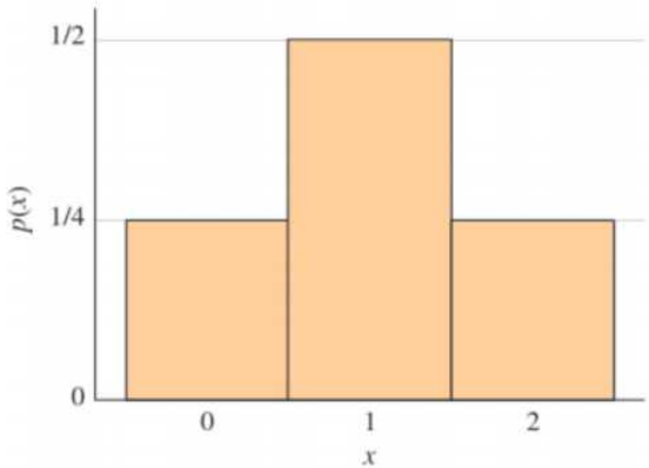
$P(x = 1) = 2/4$

$P(x = 2) = 1/4$

Probability Distribution

x	Simple Events in x	$p(x)$
0	E_4	1/4
1	E_2, E_3	1/2
2	E_1	1/4
		$\Sigma\, p(x) = 1$

Probability histogram



Probability Distributions

- Probability distributions can be used to describe the population, just as we described samples in Chapter 1.
  - **Shape**: Symmetric, skewed, mound-shaped…
  - **Outliers**: unusual or unlikely measurements
  - **Center and spread**: mean and standard deviation. A population mean is called μ and a population standard deviation is called σ.

Mean and Standard Deviation for a Discrete RV.

- Let x be a discrete random variable with probability distribution $p(x)$. Then the mean, variance and standard deviation of x are given as

Mean: $\mu = E(x) = \sum x\, p(x)$

Variance : $\sigma^2 = E[(x-\mu)^2] = \sum (x-\mu)^2\, p(x)$

$$= \sum x^2\, p(x) - \boldsymbol{\mu}^2$$

Standard deviation: $\sigma = \sqrt{\sigma^2}$

Example:

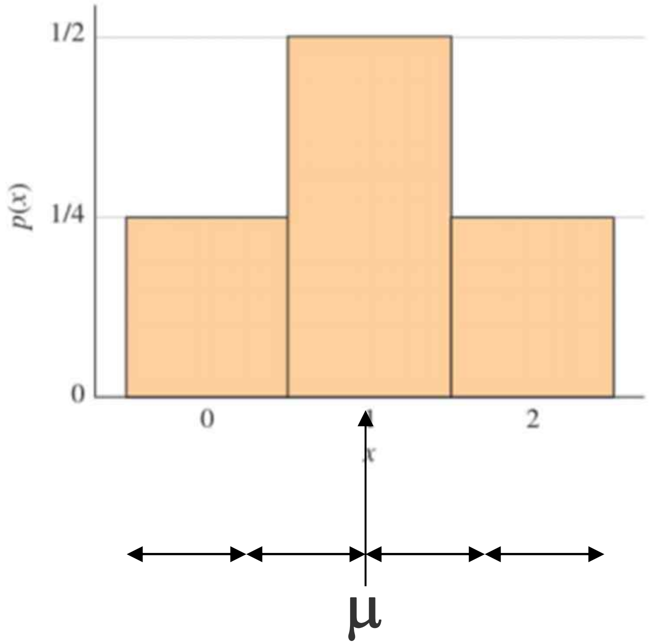
- Toss a fair coin 2 times and record *x* the number of heads.

x	$p(x)$	$xp(x)$	$(x-\mu)^2 p(x)$
0	1/4	0	$(-1.0)^2(1/4)$
1	2/4	2/4	$(0)^2(2/4)$
2	1/4	2/4	$(1.0)^2(1/4)$

$$\mu = \sum x\, p(x) = \frac{4}{4} = 1.0$$

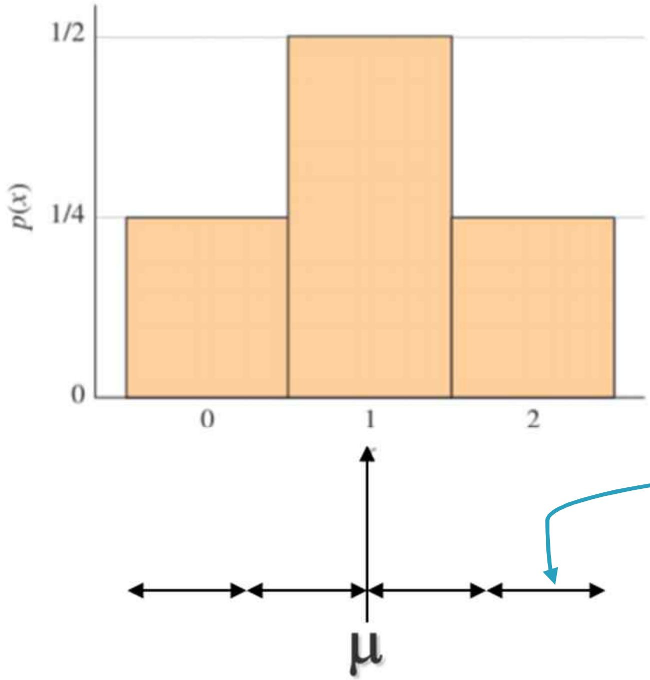
$$\sigma^2 = \sum (x-\mu)^2\, p(x) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

$$\sigma = \sqrt{0.5} = .707$$



Example:

- Toss a fair coin 2 times and record x the number of heads.



- Shape? Symmetric
- Outliers? None
- Center? $\mu = 1.5$
- Spread? $\sigma = .707$

Example

- Probability distribution of defective computer parts

x	0	1	2	3	4	5
P(x)	.02	.20	.30	.30	.10	.08

compute standard deviation of x

Sol)

x	$P(x)$	$xP(x)$	x^2	$x^2P(x)$
0	.02	.00	0	.00
1	.20	.20	1	.20
2	.30	.60	4	1.20
3	.30	.90	9	2.70
4	.10	.40	16	1.60
5	.08	.40	25	2.00
		$\Sigma xP(x) = 2.50$		$\Sigma x^2P(x) = 7.70$

$\mu = \sum \ x\,p(x)$ = 2.5

$\sigma^2 = \sum x^2\,p(x) - \mu^2 = 7.7 - 2.5^2 = 1.45 \ , \ \sigma = \sqrt{\sigma^2} = \sqrt{1.45} = 1.204$

Conditions of a Binomial Experiment

1. There are n identical trials.

2. Each trial has only two possible outcomes (or events). In other words, the outcomes of a trial are divided into two mutually exclusive events.

   - two outcomes : *success and failure*

3. The probabilities of the two outcomes (or events) remain constant.

   - $P(success) = p, \quad P(failure) = q, \quad p + q = 1$

4. The trials are independent.

Example

- Consider Ten tosses of a coin.

Determine whether or not it is a binomial experiment.

Sol)

1. There are a total of 10 trials (tosses), and they are all identical.

   - $n = 10$

2. Each trial (toss) has only two possible outcomes: a head and a tail. Let a head be called a success and a tail be called a failure.

3. $p = P(Head) = \frac{1}{2}, \quad q = P(Tail) = 1/2 \quad \Rightarrow p + q = 1$

   - these probabilities remain the same for each toss.

4. The trials (tosses) are independent.

The Binomial Probability Distribution

A binomial experiment consists of n identical trials with probability of success p on each trial.

The probability of k successes in n trials is

$$P(x = k) = C_k^n\, p^k\, q^{n-k} = \frac{n!}{k!(n-k)!}\, p^k q^{n-k}$$

for values of $k = 0, 1, 2, \dots, n$.

where $C_k^n = \frac{n!}{k!(n-k)!}$ Or use notation $C_k^n = \binom{n}{k}$

$$n! = n(n-1)(n-2)\dots(2)(1) \quad and \quad 0! = 1$$

Mean and Standard Deviation for Binomial RV.

The random variable x , the number of successes in n trials, has a probability distribution with this center and spread:

Mean: $\mu = np$

Variance: $\sigma^2 = npq$

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

Example

Find $P(x = 2)$ for a binomial random variable with $n = 10$ and $p = .1$.

Sol)

- $P(x = 2)$ is the probability of observing 2 successes and 8 failures in a sequence of 10 trials.

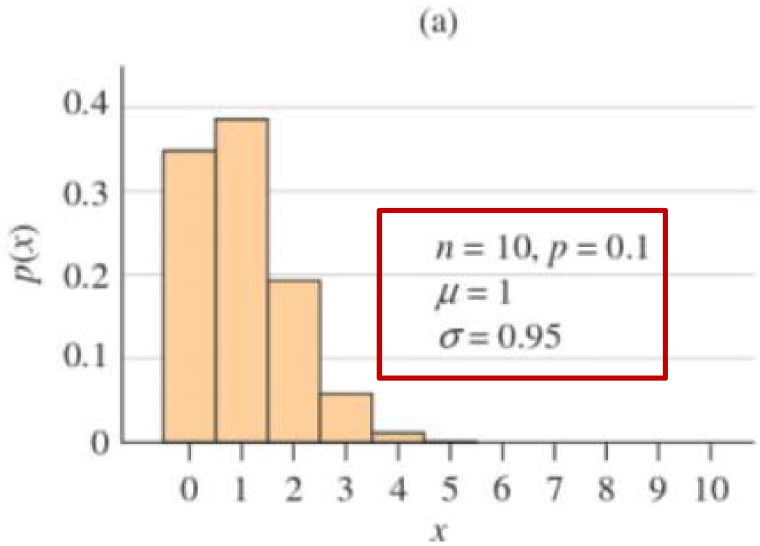
  S, S, F, F, F, F, F, F, F, F $\Rightarrow$ $ppqqqqqqqq \quad = p^2q^8$

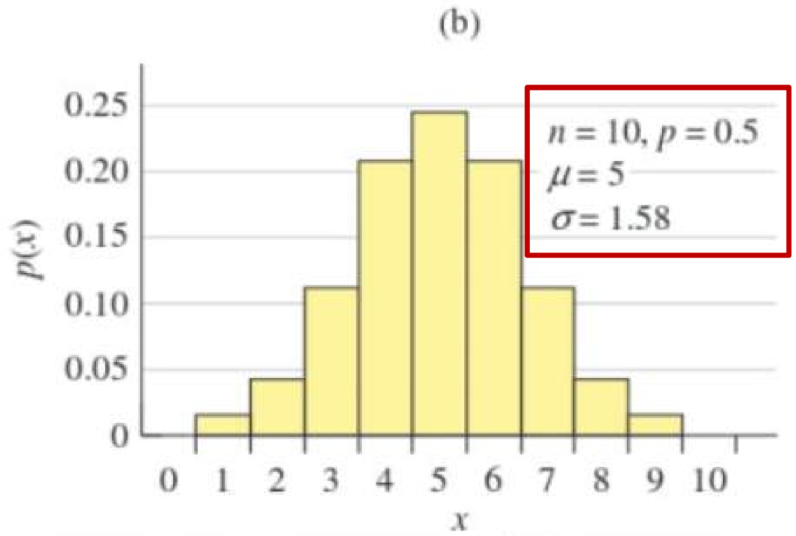
  There are C_2^{10} cases for $x = 2$ successes.

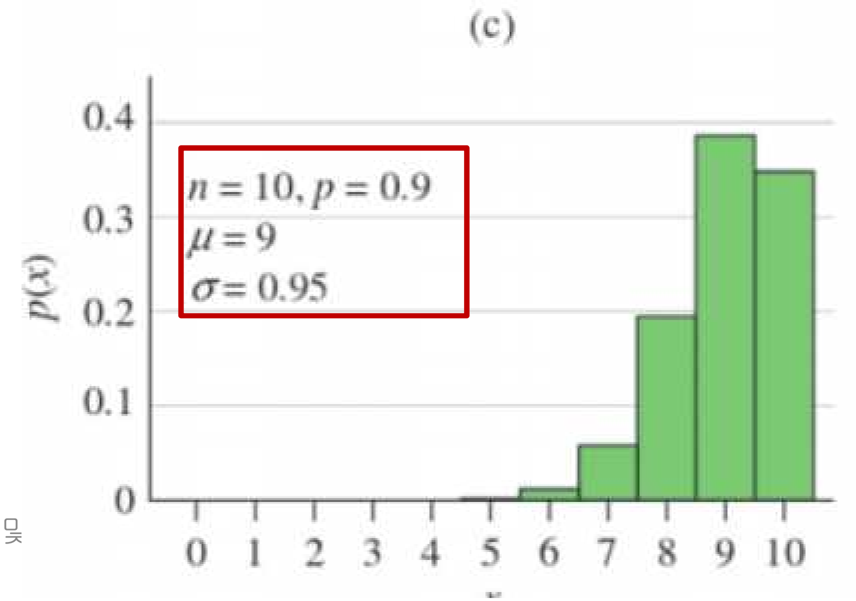
- $P(x = 2) = C_2^{10}\,(.1)^2\,(.9)^{10-2} = \frac{10!}{2!(10-2)!}\,(.1)^2(.9)^{10-2}$

  $= \frac{10(9)}{2(1)}\,(.01)(.430467) = .1937$

Binomial Probability Distributions







Cumulative Binomial Probabilities

- Calculating binomial probabilities becomes tedious even for relatively small values of n
- As n gets larger, it becomes almost impossible without the help of a calculator or computer.

⇒ Use tables of cumulative binomial probabilities are given in Table 1 of Appendix I for values of n ranging from 2 to 25 and for selected values of p.

Ex) For n=5 and p=.6 $P(x \le 3) = p(0) + p(1) + p(2) + p(3) = .663$

$$P(x = 3) = P(x \le 3) - P(x \le 2) = .663 - .317 = .346$$

$$P(x \ge 3) = 1 - P(x \le 2) = 1 - .317 = .683$$

Table 1 in Appendix I

							p							
k	.01	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	.99	k
0	—	—	—	—	—	—	—	.010	—	—	—	—	—	0
1	—	—	—	—	—	—	—	.087	—	—	—	—	—	1
2	—	—	—	—	—	—	—	.317	—	—	—	—	—	2
3	—	—	—	—	—	—	—	.663	—	—	—	—	—	3
4	—	—	—	—	—	—	—	.922	—	—	—	—	—	4
5	—	—	—	—	—	—	—	1.000	—	—	—	—	—	5

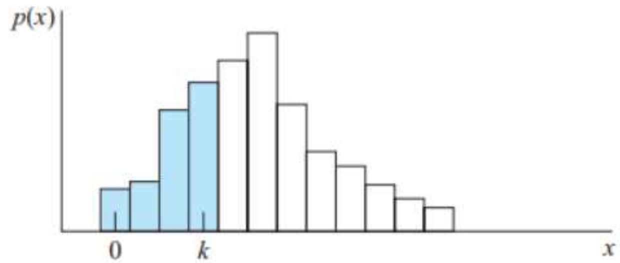


TABLE 1 Cumulative Binomial Probabilities

Tabulated values are $P(x \le k) = p(0) + p(1) + \cdots + p(k)$.
(Computations are rounded at the third decimal place.)

$n = 2$

	p													
k	.01	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	.99	k
0	.980	.902	.810	.640	.490	.360	.250	.160	.090	.040	.010	.002	.000	0
1	1.000	.998	.990	.960	.910	.840	.750	.640	.510	.360	.190	.098	.020	1
2	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	2

$n = 3$

	p													
k	.01	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	.99	k
0	.970	.857	.729	.512	.343	.216	.125	.064	.027	.008	.001	.000	.000	0
1	1.000	.993	.972	.896	.784	.648	.500	.352	.216	.104	.028	.007	.000	1
2	1.000	1.000	.999	.992	.973	.936	.875	.784	.657	.488	.271	.143	.030	2
3	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	3

n = 2, 3, …12, 15, 20, 25

Example

- Consider binomial random variable x with n=5 and p=.6
- Construct the probability histogram for the random variable x and describe its shape and location by using cumulative binomial table

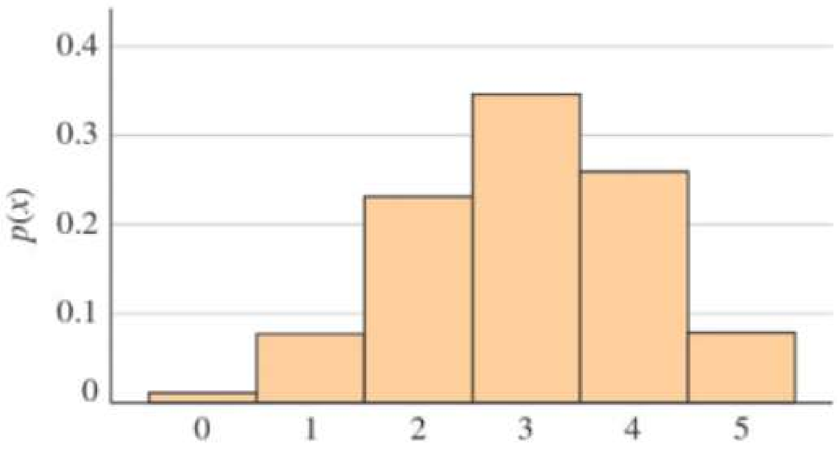
Sol) $P(x=0) = .010$ *from table*

$$P(x=1) = P(x \le 1) - P(x=0) = .087 - .010 = .077$$

$$\mathrm{P}(\mathrm{x}=2) = P(x \le 2) - P(x \le 1) = .317 - .087 = .230$$

Similarly $P(x=3) = .346,\ P(x=4) = .259,\ P(x=5) = .078$

Binomial Probability distribution



$$\mu = np = 5 * 0.6 = 3.0$$

$$\sigma^2 = npq = 5 * 0.6 * 0.4 = 1.2$$

$$\sigma = \sqrt{1.2} = 1.095$$

Distribution is relatively symmetric(mound) shaped.

The Poisson Probability Distribution

- Let x represent 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.

  - Assumption: events occur randomly and independently of one another.

$\Rightarrow$ the random variable x can be modeled by the **Poisson random variable**

- Let μ be the average number of times that an event occurs in a certain period of time or space. The probability of k occurrences of this event is

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

for values of $k = 0, 1, 2, 3, , \ldots$ $\qquad e = 2.71828$

- The mean and standard deviation of the Poisson random variable x are

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

Example : Poisson Random variables

- The number of calls received by a technical support specialist during a given period of time
- The number of bacteria per small volume of fluid
- The number of customer arrivals at a checkout counter during a given minute
- The number of machine breakdowns during a given day
- The number of traffic accidents on a section of freeway during a given time period

Example

- The average number of traffic accidents on a certain section of highway is two per week. Assume that the number of accidents follows a Poisson distribution with $\mu = 2$.
  1. Find the probability of no accidents on this section of highway during a 1-week period.
  2. Find the probability of at most three accidents on this section of highway during a 2-week period.

Sol)

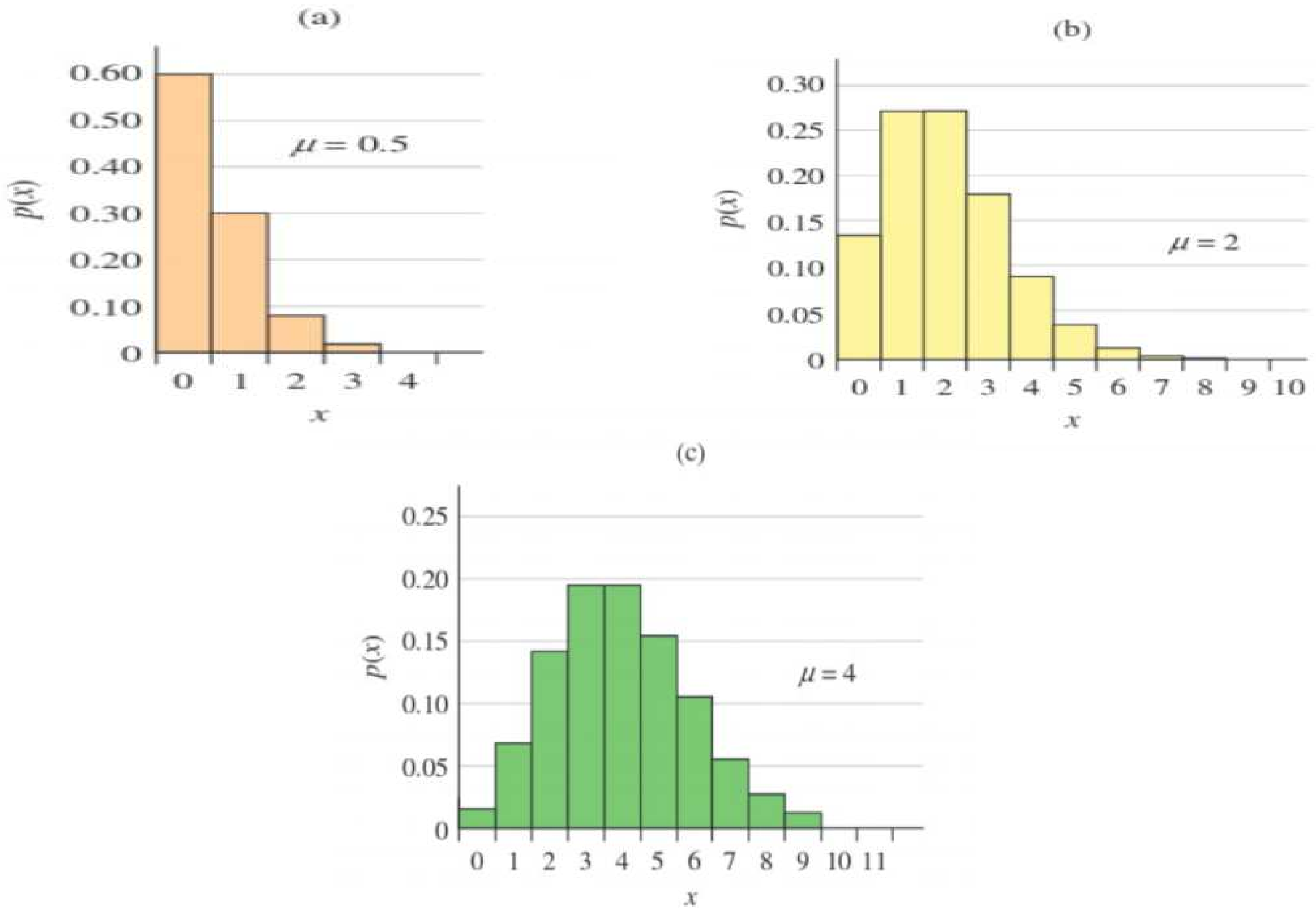
1. $k = 0$, $P(x = 0) = p(0) = \dfrac{2^0 e^{-2}}{0!} = e^{-2} = .135335$

2. 2 week period $\Rightarrow \mu = 2 \times 2 = 4$

$$P(x \leq 3) = p(0) + p(1) + p(2) + p(3)$$

$$= \frac{4^0 e^{-4}}{0!} + \frac{4^1 e^{-4}}{1!} + \frac{4^2 e^{-4}}{2!} + \frac{4^3 e^{-4}}{3!} = .433471$$

Poisson probability distributions



Cumulative Poisson Table : Table 2 in Appendix I

TABLE 2 Cumulative Poisson Probabilities

Tabulated values are $\boxed{P(x \le k) = p(0) + p(1) + \cdots + p(k).}$

(Computations are rounded at the third decimal place.)

	μ										
k	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0	1.5
0	.905	.819	.741	.670	.607	.549	.497	.449	.407	.368	.223
1	.995	.982	.963	.938	.910	.878	.844	.809	.772	.736	.558
2	1.000	.999	.996	.992	.986	.977	.966	.953	.937	.920	.809
3		1.000	1.000	.999	.998	.997	.994	.991	.987	.981	.934
4				1.000	1.000	1.000	.999	.999	.998	.996	.981
5							1.000	1.000	1.000	.999	.996
6										1.000	.999
7											1.000

$$P(x=0) = p(0) = \frac{2^0\, e^{-2}}{0!} = .135$$

	μ										
k	2.0	2.5	3.0	3.5	4.0	4.5	5.0	5.5	6.0	6.5	7.0
0	.135	.082	.050	.033	.018	.011	.007	.004	.003	.002	.001
1	.406	.287	.199	.136	.092	.061	.040	.027	.017	.011	.007
2	.677	.544	.423	.321	.238	.174	.125	.088	.062	.043	.030
3	.857	.758	.647	.537	.433	.342	.265	.202	.151	.112	.082
4	.947	.891	.815	.725	.629	.532	.440	.358	.285	.224	.173

THE POISSON APPROXIMATION TO THE BINOMIAL DISTRIBUTION

- Binomial probabilities when n is large and $\mu = np$ is small (preferably with $np < 7$)

  $\Rightarrow$ Use Poisson probability 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.

Sol)

- Exact probability : binomial dist. ($n = 5000,\ p = .001$)

$$P(x = 4) = C_4^{5000}\ (.001)^4\ (.999)^{4996}$$

- Poisson approximation: $\mu = 5000 \times .001 = 5$

$$P(x = 4) \approx \frac{\mu^k e^{-\mu}}{k!} = \frac{5^4 e^{-5}}{4!} = \frac{(625)(.006728)}{24} = .175$$ or Use Table

- The individual Poisson probabilities for $\mu = 1$ along with the individual binomial probabilities for $n = 1000,\ p = .001$ and $x = 0, 1, \ldots, 10$

x	Binomial $p(x)$	x	Poisson $p(x)$
0	0.3677	0	0.3679
1	0.3681	1	0.3679
2	0.1840	2	0.1839
3	0.0613	3	0.0613
4	0.0153	4	0.0153
5	0.0030	5	0.0031
6	0.0005	6	0.0005
7	0.0001	7	0.0001
8	0.0000	8	0.0000
9	0.0000	9	0.0000
10	0.0000	10	0.0000

✓ Very Similar

THE HYPERGEOMETRIC PROBABILITY DISTRIBUTION

- Consider a bowl containing M red balls and N-M white balls, for a total of N balls in the bowl.

  $\Rightarrow$ select n balls from the bowl and record x, the number of red balls

✓ X = the number of red ball("success")

: hypergeometric random variable

THE HYPERGEOMETRIC PROBABILITY DISTRIBUTION

- A population contains M successes and N-M failures. The probability of exactly k successes in a random sample of size n is

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

for values of k that depend on N, M, and n with $C_n^N = \frac{N!}{n!(N-n)!}$

- Mean and variance of the hypergeometric random variable:

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

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

Example

- A case of wine has 12 bottles, 3 of which contain spoiled wine.
- A sample of 4 bottles is randomly selected from the case.

1. Find the probability distribution for x, the number of bottles of spoiled wine in the sample.
2. What are the mean and variance of x?

Sol)

1. N=12, n=4, M=3, and (N-M)=9.

$$p(x) = \frac{C_x^M C_{n-x}^{N-M}}{C_n^N} = \frac{C_x^3 C_{4-x}^9}{C_4^{12}}$$

$$p(0) = \frac{C_0^3 C_4^9}{C_4^{12}} = \frac{1(126)}{495} = .25,\ p(1) = .51, p(2) = .22, p(3) = .02$$

2. Mean and variance

$$\mu = n\left(\frac{M}{N}\right) = 4\left(\frac{3}{12}\right) = 1,\ \ \sigma^2 = 4\left(\frac{3}{12}\right)\left(\frac{9}{12}\right)\left(\frac{12-4}{11}\right) = .5455$$

Key Concepts

I. Discrete Random Variables and Probability Distributions

1. Random variables, discrete and continuous

2. Properties of probability distributions

$$\mathbf{0 \le p(x) \le 1} \; \textit{and} \; \sum \boldsymbol{p}\,(\boldsymbol{x}) = \mathbf{1}$$

3. Mean or expected value of a discrete random variable:

$$\textbf{Mean: } \boldsymbol{\mu} = \sum \boldsymbol{x}\,\boldsymbol{p}(\boldsymbol{x})$$

4. Variance and standard deviation of a discrete random variable:

$$\textbf{Variance: } \boldsymbol{\sigma}^2 = \sum (\boldsymbol{x} - \boldsymbol{\mu})^2 \boldsymbol{p}(\boldsymbol{x})$$

$$\textit{Standard deviation: } \boldsymbol{\sigma} = \sqrt{\boldsymbol{\sigma}^2}$$

Key Concepts

II. The Binomial Random Variables

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

3. Mean of the binomial random variable: $\mu = np$

4. Variance and standard deviation: $\sigma^2 = npq$ *and* $\sigma = \sqrt{npq}$

Key Concepts

III. 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
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

IV. 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:

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