Introduction to Statistical Modelling

STAT2507D

Chapter 5

Discrete Distributions

INTRODUCTION

- Discrete random variables take on only a finite or countable 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

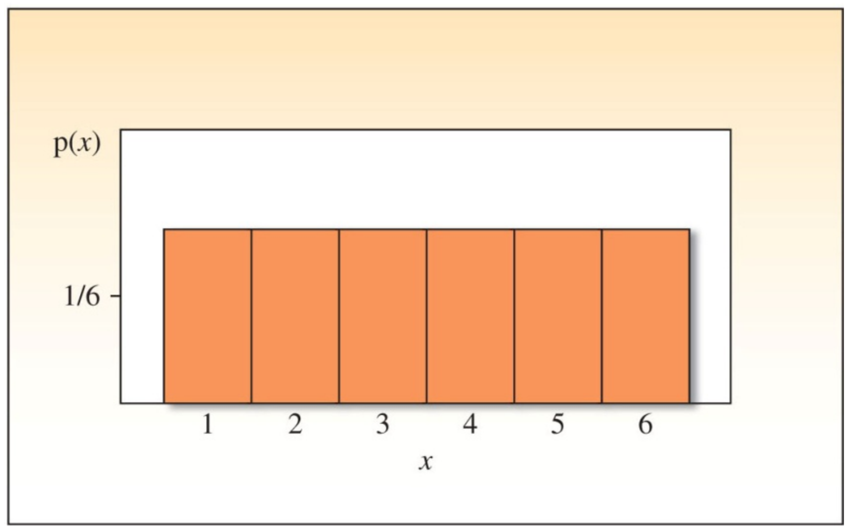
The Discrete Uniform distribution

- The probability mass function of uniform random variable X is given by

$$p(x) = \frac{1}{k}, \quad \mathrm{x} = 1, 2 \cdots, k$$

THE DISCRETE UNIFORM DISTRIBUTION

- Example: x = the number that appears when rolling a die. x has equally likely possible values. x = 1, 2, 3, 4, 5, 6 with p(x) = 1/6 for all values of x.



THE BERNOULLI DISTRIBUTION

- The Bernoulli trial is an experiment with only two possible outcomes (success– 1;failure - 0), with positive probabilities p and 1-p
- Examples: Flipping a coin, Hitting a target, Results of an exam – pass or fail
- Bernoulli random variable X,

$$\mathrm{x} = \begin{cases} 1, & if\ success\ occured \\ 0, & if\ failure\ occured \end{cases}$$

THE BERNOULLI DISTRIBUTION

- Probability of Bernoulli random variable is expressed as

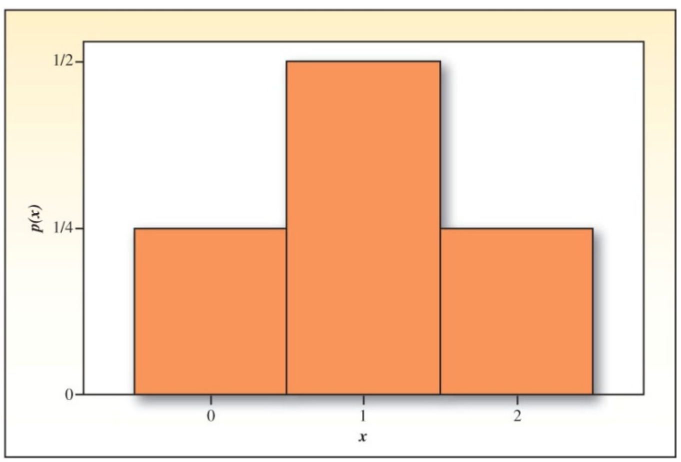
$$p(x) = p^x (1-p)^{1-x}, \quad \text{for } x = 0, 1$$

$$\text{where} \quad 0 \le p(x) \le 1, \text{ for } x = 0, 1$$

$$\sum p(x) = 1$$

THE BINOMIAL RANDOM VARIABLE

- The coin-tossing experiment is a simple example of a binomial random variable. Toss a fair coin n=2 times and record x = number of heads



x	$p(x)$
0	1/4
1	1/2
2	1/4

THE BINOMIAL RANDOM VARIABLE

- Many situations in real life resemble the coin toss, but the coin is not necessarily fair, so that $P(H) \neq 0.5$
  - Proportion of population who possess a gene linked to Alzheimer's disease = $p \neq 0.5$

THE BINOMIAL EXPERIMENT

1. 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

Example: Select two people from the Canada population, and suppose that 70% of the population has the gene for tasting phenylthiocarbomide (PTC) (N=35million, n=2)

- For the first person, p = P(gene) = 0.70
- For the second person, $p \approx$ P(gene) = 0.70, even though one person has been removed from the population

Binomial or Not

Example: A bottle of medication with 18 pills of prescribed medication and 2 pills of generic equivalent of prescribed medication (N = 20, n=2)

$$P(\text{generic on trial } 2 \mid \text{generic on trial } 1) = 1/19$$

$$\text{P}(\text{generic on trial } 2 \mid \text{not generic on trial } 1) = 2/19$$

If n/N ≥ 0.05, then the resulting experiment is not binomial

THE BINOMIAL PROBABILITY DISTRIBUTION

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

$$P(X=k)=C_k^n p^k (1-p)^{n-k}, \quad \text{for } k=1,2,\cdots,n$$

Where
$$C_k^n = \frac{n!}{k!(n-k)!}$$

$$n! = n(n-1)(n-2)\cdots(2)(1), \text{ and } \quad 0! = 1$$

The Binomial Probability Distribution

- Mean:

$$\mu = np$$

- Variance:

$$\sigma^2 = np(1 - p) = npq$$

- Standard deviation:

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

EXERCISE

5.1) Consider binomial random variable with $n = 8$ and $p = 0.7$. Let X be the number of successes in the sample. Find the following probabilities.

a) $P(x = 0)$
b) $P(x = 2)$
c) $P(x \leq 2)$

CUMULATIVE PROBABILITY TABLES

- Cumulative Probability:

  $$P(X \leq k) = P(X = 0) + \cdots + P(X = k)$$

- Find the table (next page) for n = 5 (For various values of n refer text book pages 712 – 717)
- Find the column for the correct value of p
- The row marked "k" gives the cumulative probability

EXERCISE

$n = 5$

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

CUMULATIVE PROBABILITY TABLES

- You can use the cumulative probability tables to find probabilities for selected binomial distributions (Minitab output)

Binomial with n = 10 and p = 0.5

x	P(X <= x)
0	0.00098
1	0.01074
2	0.05469
3	0.17187
4	0.37695
5	0.62305
6	0.82813
7	0.94531
8	0.98926
9	0.99902
10	1.00000

Binomial with n = 10 and p = 0.5

x	P(X = x)
0	0.000977
1	0.009766
2	0.043945
3	0.117188
4	0.205078
5	0.246094
6	0.205078
7	0.117188
8	0.043945
9	0.009766
10	0.000977

EXERCISE

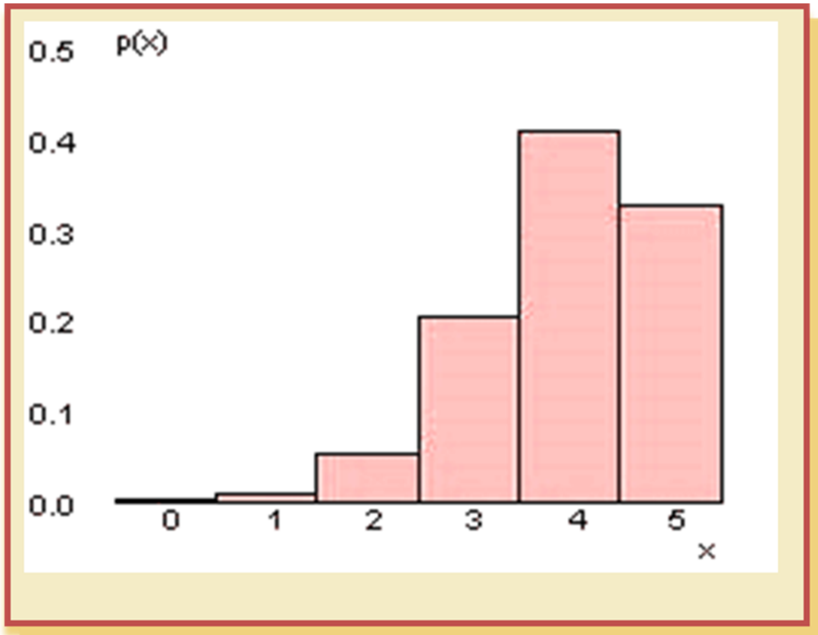
5.2) A marksman hits a target 80% of the time. He fires five shots at the target.

a) What is the probability that exactly 3 shots and more than 3 shots hit the target?

b) What are the mean, standard deviation and probability distribution for X?

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

EXERCISE



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

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

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

EXERCISE

5.3) A fair coin is tossed 4 times. Find the probability of getting exactly 2 heads.

DISCRETE PROBABILITIES WHEN USING CUMULATIVE TABLE

$$P(X < a) = P(X \le (a-1))$$

$$P(X > a) = 1 - P(X \le a)$$

$$P(X \ge a) = 1 - P(X \le (a-1))$$

$$P(a \le X \le b) = P(X \le b) - P(X \le (a-1))$$

$$P(a < X < b) = P(X \le (b-1)) - P(X \le a)$$

$$P(a < X \le b) = P(X \le b) - P(X \le a)$$

$$P(a \le X < b) = P(X \le (b-1)) - P(X \le (a-1))$$

EXERCISE

5.4) Binomial random variable with n = 8 and p = 0.7. Let X be the number of successes in the sample. Find the probability using the formula and using the cumulative table

a) $P(X \leq 3)$
b) $P(X \geq 3)$
c) $P(X < 3)$
d) $P(X = 3)$
e) $P(2 \leq X \leq 5)$

EXERCISE

5.5)A home security system is designed to have a 99% reliability rate. Suppose that nine homes equipped with this system experience an attempted burglary. Find the probability of these events: At least one of the alarms is triggered; More than 7 of them are triggered; 8 or fewer alarms are triggered.

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. Events occur randomly and independently of one another.

POISSON RANDOM VARIABLES

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

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:

$$P(X = k) = \frac{\mu^k e^{-\mu}}{k!}, \quad k = 0, 1, 2, \cdots$$

THE POISSON PROBABILITY DISTRIBUTION

- Mean and Variance for Poisson Distribution
  - Mean: $\mu = \mu$
  - Variance: $\sigma^2 = \mu$

EXERCISE

5.6) Consider a Poisson random variable X with $\mu = 3$. Use the Poisson formula to calculate the following probabilities

a) $P(x=0)$

b) $P(x=1)$

c) $P(x>1)$

d) $P(x \leq 2)$

CUMULATIVE PROBABILITY TABLES

- Using cumulative probability tables to find probabilities for selected Poisson distributions (page718-719)
- Find the column for the correct value of mean. The row marked “k” gives the cumulative probability, $P(X \leq k) = P(X = 0) + \ldots + P(X = k)$

CUMULATIVE PROBABILITY TABLES

- Poisson probability table

k	μ = 2.0	2.5	3.0	3.5	4.0
0	0.135	0.082	0.055	0.033	0.018
1	0.406	0.287	0.199	0.136	0.092
2	0.677	0.544	0.423	0.321	0.238
3	0.857	0.758	0.647	0.537	0.433
4	0.947	0.891	0.815	0.725	0.629
5	0.983	0.958	0.916	0.858	0.785
6	0.995	0.986	0.966	0.935	0.889
7	0.999	0.996	0.988	0.973	0.949
8	1.000	0.999	0.996	0.990	0.979
9		1.000	0.999	0.997	0.992
10			1.000	0.999	0.997
11				1.000	0.999
12					1.000

EXERCISE

5.7) Consider a Poisson random variable X with $\mu = 3$. Use Poisson cumulative probability table to find the following

a) $P(x \leq 3)$
b) $P(x > 3)$
c) $P(x = 3)$
d) $P(3 \leq x \leq 5)$

EXERCISE

5.8) The average number of traffic accidents on a certain section of highway is two per week. Find the probability of exactly one accident and 8 or more during a one-week period.

EXAMPLE

5.9) Bank customers arrive on weekday afternoons at an average of 3.0 customers every 4 minutes

a. What is the probability of having more than 7 customers in a 4-minute interval on a weekday afternoon?
b. What is the probability of getting exactly 10 customers during an 8-minute interval on a weekday afternoon?

THE POISSON APPROXIMATION TO BINOMIAL

- The Poisson probability distribution provides a simple and accurate approximation to binomial probabilities when n is large and $\mu = np$ is small, preferably with $np < 7$.

EXERCISE

5.10)To illustrate how well the Poisson probability distribution approximates the binomial probability distribution, calculate the Poisson approximate values for p(0) and p(1) for a binomial probability distribution with n=25 and p = 0.05. Compare the answers with the exact values obtained from Table 1 in Appendix I

THE HYPERGEOMETRIC PROBABILITY DIST'N

- Suppose that a certain population consists of N items in total. Out of those N items, M have a specific characteristic that is of interest (successes), and hence, (N –M) do not possess this characteristic (failures). If we randomly select a sample of size n, with no replacement, the random variable, X, that will represent the number of items obtained from those that have characteristic M, follows a hypergeometric distribution

THE HYPERGEOMETRIC PROBABILITY DIST'N

- The probability of exactly k successes in n trials is

$$P(X = k) = \frac{\mathrm{C}_{\mathrm{k}}^{\mathrm{M}}\mathrm{C}_{\mathrm{n-k}}^{\mathrm{N-M}}}{\mathrm{C}_{\mathrm{n}}^{\mathrm{N}}}$$

- The values of k depend on N, M and n
- Mean: $\mu = n\left[\frac{M}{N}\right]$
- Variance: $\sigma^2 = n\left[\frac{M}{N}\right]\left[\frac{N-M}{N}\right]\left[\frac{N-n}{N-1}\right]$

THE HYPERGEOMETRIC PROBABILITY DIST'N

- The "M&M" problems from chapter 4 (Exercise 4.12) 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 M&M" to a 'success'. Probability of k red candies is

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

- If the selection is done with replacement, the distribution becomes binomial

EXERCISE

5.11) Let X be the number of successes observed in a sample of n=5 items selected from N=10. Suppose that, of the N=10 items, 6 are considered “successes”.

a) Find the probability of observing no successes
b) Find the probability of observing at least 2 successes
c) Find the probability of observing exactly 2 successes

EXERCISE

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

SUMMARY

- The Binomial Random Variable
  - 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
    - x is the number of successes in n trials

SUMMARY

- Calculating binomial probabilities
  - Formula: $P(x = k) = C_k^n p^k (1 - p)^{n-k}, for\ k = 0,1, \cdots n$
  - Cumulative binomial tables
  - Individual and cumulative probabilities using Minitab

SUMMARY

- The Binomial Random Variable cont'd
  - Mean, variance and std of binomial random variable
    - Mean: $\mu = np$
    - Variance: $\sigma^2 = npq$
    - Standard deviation: $\sigma = \sqrt{npq}$

SUMMARY

- The Poisson Random Variable
  - 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
  - Calculating Poisson probabilities
    - Formula: $P(x = k) = \frac{e^{-\mu}\mu^{k}}{k!}$, for k=0,1, ….
    - Cumulative Poisson Tables
    - Individual and cumulative probabilities using Minitab

SUMMARY

- The Poisson Random Variable cont'd
  - Mean, Variance, Std of Poisson random variable
    - Mean: $\mu = \mu$
    - Variance: $\sigma^2 = \mu$
    - Standard deviation: $\sigma = \sqrt{(\mu)}$
  - Binomial probabilities can be approximated with Poisson probabilities when $np < 7$, using $\mu = np$

SUMMARY

- The Hypergeometric Random Variable
  - The number of successes in a sample of size n from a finite population containing M successes and N-M failures
  - Formula for the probability of k successes in n trials: $P(x = k) = \frac{C_k^M C_{n-k}^{N-M}}{C_n^N}$

SUMMARY

- The Hypergeometric Random Variable cont'd
  - Mean, variance and std of the hypergeometric random variable
    - Mean: $\mu = n\left[\frac{M}{N}\right]$
    - Variance: $\sigma^2 = n\left[\frac{M}{N}\right]\left[\frac{N-M}{N}\right]\left[\frac{N-n}{N-1}\right]$
    - Standard deviation: $\sigma = \sqrt{\left[\frac{M}{N}\right]\left[\frac{N-M}{N}\right]\left[\frac{N-n}{N-1}\right]}$