DISCRETE PROBABILITY DISTRIBUTIONS

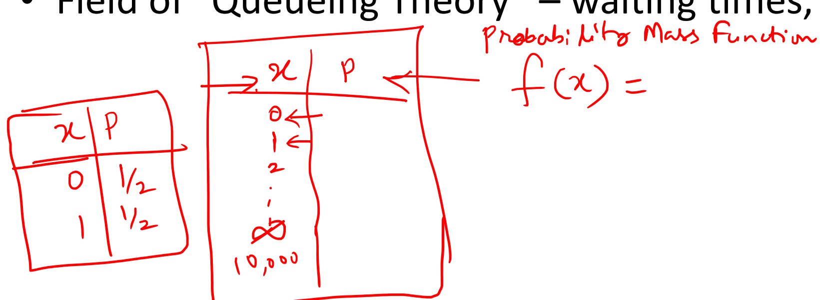
B. Srirangacharyulu

1

Discrete Probability Distributions

Introduction

- Number of voters (arrivals) at a polling station
  - Follows a probability distribution (Poisson)
- x = Number of arrivals per minute (unit time)
- Field of "Queueing Theory" – waiting times, etc.



2

Discrete Probability Distributions

Introduction

- x = Number of arrivals per minute (unit time)

x	Probability
0	0.1353
1	0.2707
2	0.2707
3	0.1804
4	0.0902
≥ 5	0.0527
TOTAL	1.000

Introduction

- Random Variables and Probability Distributions are models for population data
- Random Variables – represent the data values
- Prob. Distribution – provides either
  - the prob of each data value, or
  - a rule for computing the prob of each data value or a set of data values

Introduction

- Discrete Probability Distributions
  - Assign probabilities based on rules (shown in a Table: one column – values of the random variable; a second column – for the associated probabilities), or
  - Compute probabilities using a mathematical function

Random Variables

- Random Variable – a numerical description of the outcome of an experiment (we cannot predict the outcome accurately)
- Discrete or Continuous
- Each possible experimental outcome - a numerical value is assigned (R.V.)
- We can't predict an instance of a R.V.

Discrete Random Variable

- A random variable (x) that may assume either
  - A finite number of values, or
  - An infinite sequence of values such as 0,1,2,…
- The values can be listed
- Score of a batsman $[x = \{0,1,2,3, \dots .\}]$, number of goals scored in a foot ball match, number of vehicles arriving at a toll booth during a time period
- Outcome can be described naturally

Discrete Random Variable

- Outcome cannot be described naturally (could be like a binary variable)
- Inspection of a part – good part or bad part
- x = 0; may correspond to Good part
- x = 1; may correspond to Bad part

- We can still assign a numerical value, but it cannot explain the outcome naturally

Discrete Random Variable

Random Experiment	Random Variable (x)	Possible values for x
Flip a coin	Face of coin showing	1 if heads; 0 if tails
Roll a die	Number of dots showing on top of die	1,2,3,4,5,6
Contact 5 customers	Number of customers who buy the product	0,1,2,3,4,5
Operate a tollbooth for 1 day	Number of vehicles that arrive	0,1,2,3,….
Offer a customer to choose 2 products	Product chosen by the customer	0 if none 1 if product A 2 if product B

Continuous Random Variable

- May assume any numerical value in an interval or collection of intervals
- Time, weight, distance, etc.
- Time between 2 successive arrivals: $x \geq 0$

Random Experiment	Random Variable (x)	Possible values for x
Fill a soft drink bottle (250ml)	Number of ml	$0 \leq x \leq 250$
Invest \$10,000 in stock market	Value of investment after one year	$x \geq 0$

Developing Probability Distributions

- x : a discrete random variable
- $f(x)$: probability function
  - Provides the probability for each value of the R.V.

- Classical, relative frequency, subjective methods – to develop tabular probability distributions

Developing Probability Distributions

- Rolling a die - Classical

Classical

x	**Probability** $f(x)$
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6
TOTAL	1.000

Developing Probability Distributions

- Automobile sales in the past 300 days – relative frequency

x	Probability $f(x)$	Frequency
0	0.18	54
1	0.39	117
2	0.24	72
3	0.14	42
4	0.04	12
5	0.01	3
TOTAL	1.000	300

Also called, empirical probability distribution

13

Developing Probability Distributions

- Importance: to compute easily the probability of a variety of events

- Q. what is the probability of 3 or more automobiles being sold during a day?

- $$= f(3) + f(4) + f(5) = 0.14 + 0.04 + 0.01 = 0.19$$

14

Developing Probability Distributions

2 conditions

$$f(x) \geq 0$$

$$\sum_{\forall x} f(x) = 1$$

Developing Probability Distributions

- A formula is often used for $f(x)$
- Rolling a die – discrete uniform probability distribution

$$f(x) = \frac{1}{n}$$

where

n is the number of values x may assume

Expected Value and Variance

- Expected value – is a measure of central location for the R.V.
- When we don't have data, but only probability is known – we don't calculate mean, but we calculate 'expected value'
- Expected value is the weighted average of the values of the R.V. where the weights are the probabilities

$$E(x) = \mu = \sum_{\forall x} x.f(x)$$

17

Expected Value and Variance

- Automobile sales in the past 300 days – relative frequency

x	Probability $f(x)$
0	0.18
1	0.39
2	0.24
3	0.14
4	0.04
5	0.01
TOTAL	1.000

Expected mean

$$E(x) = \mu = \sum_{\forall x} x.f(x)$$

$$= (0 \times 0.18) + (1 \times 0.39) + (2 \times 0.24) + (3 \times 0.14) + (4 \times 0.04) + (5 \times 0.01) = 1.5 \; automobiles\ per\ day$$

$\frac{w_1 x_1 + w_2 x_2 + \dots}{w_1 + w_2 + \dots}$

$\frac{\sum x f(x)}{\sum f(x) = 1}$

18

Expected Value and Variance

- Variance – is a measure of variability in the values of the R.V.

$$Var(x) = \sigma = \sum_{\forall x} (x-\mu)^2 . f(x)$$

$$= \sum_{\forall x} \left(x - E(x)\right)^2 . f(x)$$

$$\sigma = \sqrt{\frac{(x-\mu)^2}{N}}$$

Expected Value and Variance

- Automobile sales in the past 300 days – relative frequency

x	**Probability** $f(x)$
0	0.18
1	0.39
2	0.24
3	0.14
4	0.04
5	0.01
TOTAL	1.000

$$Var(x) = \sigma^2$$

$$= \sum_{\forall x} \left(x - E(x)\right)^2 . f(x)$$

$$= [(0-1.5)^2 \times 0.18]$$
$$+ [(1-1.5)^2 \times 0.39]$$
$$+ [(2-1.5)^2 \times 0.24]$$
$$+ [(3-1.5)^2 \times 0.14]$$
$$+ [(4-1.5)^2 \times 0.04]$$
$$+ [(5-1.5)^2 \times 0.01]$$
$$= 1.25 \; automobiles$$

Bivariate Distributions

- The probability distribution involves TWO random variables
- Rolling a pair of dice
  - Outcome: number with the first die + number with the second die – (x + y)
- We are interested in the relationship between x and y
- Covariance and correlation coefficient to measure the linear association

Bivariate Distributions

- Number of automobiles sold over 300 days

Geneva Dealership	**Saratoga Dealership**						
	0	1	2	3	4	5	Total
0	21	30	24	9	2	0	86
1	21	36	33	18	2	1	111
2	9	42	9	12	3	2	77
3	3	9	6	5	5	0	26
Total	54	117	72	42	12	3	300 days

Bivariate Distributions

- Bivariate empirical probability distribution

Geneva Dealership	Saratoga Dealership						
	0	1	2	3	4	5	Total
0	0.0700	0.1000	0.0800	0.0300	0.0067	0.0000	0.2867
1	0.0700	0.1200	0.1100	0.0600	0.0067	0.0033	0.3700
2	0.0300	0.1400	0.0300	0.0400	0.0100	0.0067	0.2567
3	0.0100	0.0300	0.0200	0.0100	0.0167	0.0000	0.0867
	0.18	0.39	0.24	0.14	0.04	0.01	1.00

Bivariate Distributions

- Expected value and variance of a linear combination of two random variables

$$E(ax + by) = aE(x) + bE(y)$$

$$Var(ax + by) \\ = a^2Var(x) + b^2Var(y) + 2ab\sigma_{xy}$$

Bivariate Distributions

Expected Value and Variance of Daily Automobile Sales
at Geneva Dealership

Outcome (x)	Frequency	$f(x)$	$x.f(x)$	$x - E(x)$	$[x - E(x)]^2$	$[x - E(x)]^2.f(x)$
0	86	0.2867	0.0000	-1.1433	1.3072	0.3747
1	111	0.3700	0.3700	-0.1433	0.0205	0.0076
2	77	0.2567	0.5133	0.8567	0.7339	0.1884
3	26	0.0867	0.2600	1.8567	3.4472	0.2988
	300		1.1433			0.8695
			$E(x)$			$Var(x)$

Bivariate Distributions

Expected Value and Variance of Daily Automobile Sales
at Saratoga Dealership

Outcome (y)	Frequency	$f(y)$	$y.f(y)$	$y - E(y)$	$[y - E(y)]^2$	$[y - E(y)]^2.f(y)$
0	54	0.1800	0.00	-1.50	2.25	0.4050
1	117	0.3900	0.39	-0.50	0.25	0.0975
2	72	0.2400	0.48	0.50	0.25	0.0600
3	42	0.1400	0.42	1.50	2.25	0.3150
4	12	0.0400	0.16	2.50	6.25	0.2500
5	3	0.0100	0.05	3.50	12.25	0.1225
	300		1.50			1.2500
			$E(y)$			$Var(y)$

Bivariate Distributions

- Covariance

$$\sigma_{xy} = \sum_{\forall i,j} [x_i - E(x_i)][y_j - E(y_j)].f(x_i, y_j)$$

Bivariate Distributions

$[x_i - E(x_i)][y_j - E(y_j)]$

Geneva Dealership (x)	Saratoga Dealership (y)					
	0	1	2	3	4	5
0	1.7150	0.5717	-0.5717	-1.7150	-2.8583	-4.0017
1	0.2150	0.0717	-0.0717	-0.2150	-0.3583	-0.5017
2	-1.2850	-0.4283	0.4283	1.2850	2.1417	2.9983
3	-2.7850	-0.9283	0.9283	2.7850	4.6417	6.4983

Bivariate Distributions

$[x_i - E(x_i)][y_j - E(y_j)].f(x_i, y_j)$

Geneva Dealership (x)	Saratoga Dealership (y) 0	1	2	3	4	5
0	0.12005	0.05717	-0.0457	-0.0515	-0.0191	0
1	0.01505	0.0086	-0.0079	-0.0129	-0.0024	-0.0017
2	-0.0386	-0.06	0.01285	0.0514	0.02142	0.01999
3	-0.0279	-0.0279	0.01857	0.02785	0.07736	0

$$\sigma_{xy} = \sum_{\forall i,j} [x_i - E(x_i)][y_j - E(y_j)].f(x_i, y_j) = 0.1350$$

Bivariate Distributions

- Correlation between random variables
- Near +1: strong positive linear relationship
- Near -1: strong negative linear relationship
- Near 0: a lack of linear relationship

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x.\sigma_y} = \frac{0.1350}{\sqrt{0.8696}.\sqrt{1.25}} = 0.1295$$

- Indicates that there is a weak positive linear relationship between x and y

Bivariate Distributions

- Expected value and variance of a linear combination of two random variables

$$E(ax + by) = aE(x) + bE(y)$$

$$Var(ax + by)$$
$$= a^2Var(x) + b^2Var(y) + 2ab\sigma_{xy}$$

Bivariate Distributions

Expected Value and Variance of Daily Total Automobile Sales

s		$f(s)$	$s.f(s)$	$s - E(s)$	$[s - E(s)]^2$	$[s - E(s)]^2.f(s)$
0	(0,0)	0.0700	0.0000	-2.6433	6.9872	0.4891
1	(0,1),(1,0)	0.1700	0.1700	-1.6433	2.7005	0.4591
2	(0,2),(1,1),(2,0)	0.2300	0.4600	-0.6433	0.4139	0.0952
3	(0,3),(1,2),(2,1),(3,0)	0.2900	0.8700	0.3567	0.1272	0.0369
4	(0,4),(1,3),(2,2),(3,1)	0.1267	0.5067	1.3567	1.8405	0.2331
5	(0,5),(1,4),(2,3),(3,2)	0.0667	0.3333	2.3567	5.5539	0.3703
6	(1,5),(2,4),(3,3)	0.0233	0.1400	3.3567	11.2672	0.2629
7	(2,5),(3,4)	0.0233	0.1633	4.3567	18.9805	0.4429
8	(3,5)	0.0000	0.0000	5.3567	28.6939	0.0000
			2.6433			2.3895
			$E(s)$			$Var(s)$

Bivariate Distributions

$$E(ax + by) = aE(x) + bE(y)$$

$$Var(ax + by) = a^2Var(x) + b^2Var(y) + 2ab\sigma_{xy}$$

$$E(s) = 2.6433$$

$$Var(s) = 2.3895$$

$$\sigma_{xy} = \frac{1}{2}[2.3895 - 0.8695 - 1.2500] = 0.1350$$

Binomial Probability Distribution

1. The experiment consists of a sequence of n identical trials
2. TWO outcomes are possible on each trial – refer one outcome as 'success' and the other as 'failure'
3. The probability of a success, denoted by p, does not change from trial to trial; same is with the probability of failure $(1 - p)$
4. The trials are independent

Binomial Probability Distribution

- 2, 3, 4 together constitute the Bernoulli process
- 1,2,3,4 together constitute the Bernoulli experiment
- Tossing a coin – 8 trials: is a Bernoulli experiment
- $x =$ the number of successes in n trials of a Bernoulli experiment

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

Binomial Probability Distribution

- Martin Clothing Store Problem: what is the probability that two out of the next three customers will make a purchase?
- $p = 0.3$ – based on experience

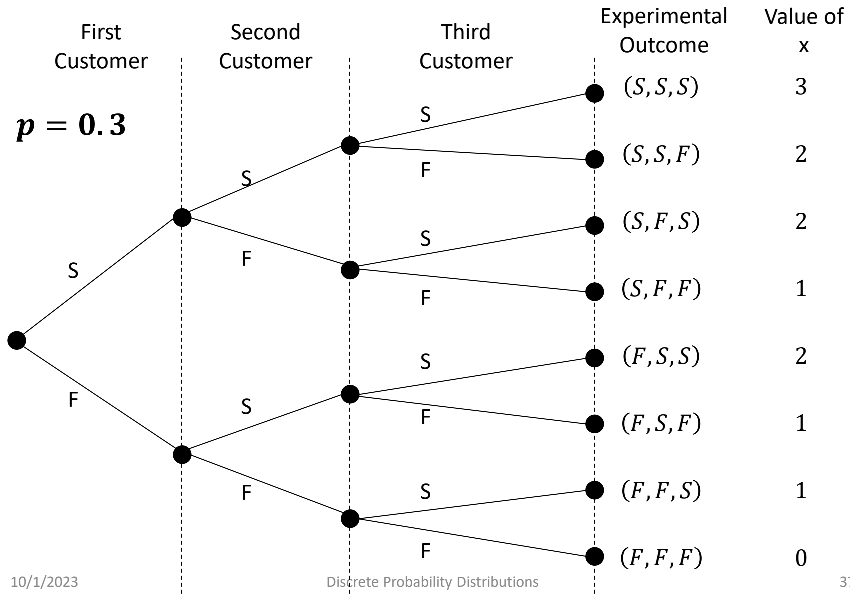
Success
Failure

$$f(x=2) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$= \binom{3}{2} (0.3)^2 (1-0.3)^{3-2} = 0.189$$

BINOM.DIST( )

Binomial Probability Distribution



Binomial Probability Distribution

A_i	Probability
S,S,S	0.3x0.3x0.3 = 0.027
S,S,F	0.063
S,F,S	0.063
S,F,F	0.147
F,S,S	0.063
F,S,F	0.147
F,F,S	0.147
F,F,F	0.343
	1.000

Binomial Probability Distribution

- Probability of 2 successes in 3 trials
  $= 0.063 + 0.063 + 0.063 = 0.189$
  $= 0.3 \times 0.3 \times 0.7$
  $= p \times p \times (1 - p)$
  $= p^x . (1 - p)^{(n-x)}$

$$\binom{n}{x} = \binom{3}{2} = 3C_2 = 3$$

A_i	Probability
S,S,S	0.3x0.3x0.3 = 0.027
S,S,F	0.063
S,F,S	0.063
S,F,F	0.147
F,S,S	0.063
F,S,F	0.147
F,F,S	0.147
F,F,F	0.343
	1.000

Binomial Probability Distribution

- Expected Value

$$E(x) = np$$

- Variance

$$Var(x) = np(1 - p)$$

Poisson Probability Distribution

- The discrete random variable is used in estimating the number of occurrences over a specified interval of time or space
- E.g. number of arrivals at a car garage in one hour, number of repairs needed in 10km of highway, number of leaks in 100 miles of pipeline
- This distribution is often used in waiting line situations

Poisson Probability Distribution

- Properties:

1. The probability of occurrence is the same for any two intervals of equal length
2. The occurrence or non-occurrence in any interval is independent of the occurrence or non-occurrence in any other interval

Poisson Probability Distribution

- Probability Distribution Function:
- $x = 0,1,2, \ldots\ldots$ the number of occurrences in an interval
- $f(x)$: The probability of x occurrences in an interval
- $e = 2.71828$
- $\mu =$ expected value or mean number of occurrences in an interval

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

Poisson Probability Distribution

- We are interested in the number of patients who arrive at the emergency room of a hospital during a 15-minute period on weekday mornings
- $\mu = 10$
- Probability of 5 arrivals:

$$f(5) = \frac{\mu^x e^{-\mu}}{x!} = \frac{10^5 e^{-10}}{5!} = 0.0378$$

- Mean and variance are equal

References

- Anderson, D.R., D.J. Sweeney, T.A. Williams, J.D. Camm, J.J. Cochran, M.J. Fry, and J.W. Ohlmann, Statistics for Business and Economics, 14th edition, Cengage Learning
- Levin, R.I., and D.S. Rubin, Statistics for Management, 7th edition, PHI