5 Lecture 6: Discrete Distributions

Wish you would learn to love people and use things, and not the other way around. Aubrey Graham
Friday's Session

- Fri Apr 21 2017
- 10:40AM - 1:30PM
- Space Assignment(s):Rachel Carson Acad 252

5.1 Random Variables

A **random variable** is a variable (denoted by X or x) that has a single event, determined by chance, can be any outcome of interest in the sample space.

- X Random variable
- x Observed variable

Example 1: Rolling a die, the outcomes can be $1, 2, 3, 4, 5, 6$. X can be any of these values.

Example 2: Tossing a coin, the outcomes can be H or T. X can be H or T.

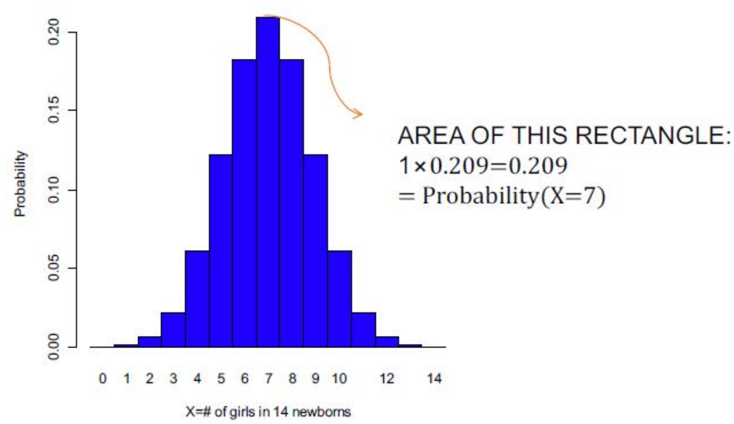
All events in the sample space has an associated probability.

Example 3: Rolling a die.

X	$P(X)$
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

This table is a **probability distribution** which assigns a probability to each of the random variables possible outcomes. This can also be used within a graph or a formula.

Example 4: The following is a probability distribution for number of new born girls.



What does the x-axis represent?

What does the y-axis represent?

5.2 Discrete and Continuous Distributions.

These random variables can be one of two types of variables.

1. Discrete Variables - has either a finite number of values or a countable number of values (countable means, an event that can be counted, can be infinity)
    - number of dogs owned
    - number of friends on Facebook
    - number of texts a day (countable)
2. Continuous Variables - has infinitely many values, and those values can be associated with measurements on continuous scale without gaps
    - GPA
    - velocity of a car
    - weight

Two Requirements for a Probability Distribution

1. $\sum P(X) = 1$ X is all values in the sample space
2. $0 \leq P(X) \leq 1$ probability of an event X will alway be between 0 and 1 (the probability of X can equal 0 or 1)

Important. Important. Important.
Finding the mean, variance, and standard deviation from a distribution is done differently.

Mean, Expected Value

$$E = \mu = \sum XP(X)$$

Variance

$$\begin{aligned}\sigma^2 &= \sum[(X-\mu)^2 P(X)] \\ &= \sum[(X^2 P(X))] - \mu^2\end{aligned}$$

Standard Deviation

$$\sigma = \sqrt{\sum[(X^2 P(X))] - \mu^2}$$

Review of **Outliers** You have outliers or unusual values if a value goes beyond

1. Maximum Value: $\mu + 3\sigma$
2. Minimum Value: $\mu - 3\sigma$

Example 5: X is the number of girls from 14 babies

X	$P(X)$	$XP(X)$	X^2	$X^2P(X)$
0	0.000	0.000	0	0.000
1	0.001	0.001	1	0.001
2	0.006	0.012	2	0.024
3	0.022	0.066	9	0.198
4	0.061	0.244	16	0.976
5	0.122	0.610	25	3.050
6	0.183	1.098	36	6.588
7	0.209	1.463	49	10.241
8	0.183	1.464	64	11.712
9	0.122	1.098	81	9.882
10	0.061	0.610	100	6.100
11	0.022	0.242	121	2.662
12	0.006	0.072	144	0.864
13	0.001	0.013	169	0.169
14	0.000	0.000	196	0.000

Find μ, σ^2, and σ. Procedure:

1. Find $\mu = \sum XP(X) = 6.993$
2. Find $\sum X^2P(X) = 52.467$

Mean, Expected Value

$$\mu = \sum XP(X) = 6.993 \approx 7$$

Which value for X has the highest probability?

It is expected to have 7 girls among 14 newborn babies.

Variance

$$\sigma^2 = \sum 52.467 - 6.993^2 = 3.564951 \approx 3.6 \text{ girls}^2$$

Standard Deviation

$$\sigma = \sqrt{\sum[(X^2P(X))] - \mu^2} = \sqrt{3.564951} = 1.9 \text{ girls}$$

Usual Values Maximum usual value: $\mu + 3\sigma = 7.0 + 3(1.9) = 12.7$ girls

Minimum usual value: $\mu + 3\sigma = 7.0 - 3(1.9) = 1.3$ girls

Extreme Values

For 14 randomly selected babies, the number of girls usually falls between 1.3 and 12.7. The probability of unusual events $P($ 13 or more girls$) = P(X \geq 13) = P(X = 13) + P(X = 14)$. This is $0.001 + 0.000 = 0.001$ (LOW value). This implies it is unusual to get 13 girls or more. This event would not happen by chance.

5.3 The Binomial Distribution

Introduction:
Tossing one coin follows a Bernoulli Distribution
The Random Variable X is Heads $X = 1$ or Tails $X = 0$
$P(X = \text{ Heads}) = P(X = 1) = 1/2$ $P(X = \text{ Tails}) = P(X = 0) = 1/2$

Binomial Distribution: Requirements

- Suppose a fixed number of trials (Ex. Flip a coin a n of times)
- The trials must be independent (Ex. flips do not affect each other)
- Each trial must have all outcomes classified into 2 categories (Ex. Tail or Head, disjoint)
- The probabilities must remain constant for each trial (Ex. P(head)=1/2, and this does not change)

Random Variable: X Meaning of X: Number of successes in n trials

Examples 6:

1. Getting 6 heads in 10 tosses of a coin. In a fair coin probability of head is 1/2.
2. Getting 3 correct answers in a multiple choice 5 question exam (student is unprepared. Each question has 5 possibilities (a,b,c,d,e)). Probability of getting a correct answer at any question is 1/5.
3. Hospital records show that of patients suffering from a certain disease, 75% die of it. Of 6 randomly selected patients, 4 will recover. P(recovering)= 1 – P(No recovering)= 1-0.75=0.25

Notation:

1. n = Number of trials
2. X = number of successes in n trials
3. p = Denotes the probability of success
4. q = Probability of failure $= 1 - p$

Note: Success does not necessarily mean something good!!!!!

Examples 6 CONTD:

1. Getting 6 heads in 10 tosses of a coin. In a fair coin probability of head is 1/2.
   - $n = 10$
   - $X = 6$
   - $p = 0.5$
   - $q = 0.5$
2. Getting 3 correct answers in a multiple choice 5 question exam (student is unprepared. Each question has 5 possibilities (a,b,c,d,e)). Probability of getting a correct answer at any question is 1/5.
   - $n = 5$
   - $X = 3$
   - $p = 0.2$
   - $q = 0.8$
3. Hospital records show that of patients suffering from a certain disease, 75% die of it. Of 6 randomly selected patients, 4 will recover. P(recovering)= 1 − P(No recovering)= 1-0.75=0.25
   - $n = 6$
   - $X = 4$
   - $p = 0.25$
   - $q = 0.75$

To find probabilities we must use the **binomial probability distribution**, which can be seen as

$$P(X = x) = \frac{n!}{(n-x)!x!} p^x q^{(n-x)} \tag{12}$$

where $x = 0, 1, 2, ..., n$

$$\binom{n}{x} = \frac{n!}{(n-x)!x!}$$

Examples 6 CONTD:
Getting 6 heads in 10 tosses of a coin. In a fair coin probability of head is 1/2. P(6 heads from 10 flips of a coin)

$$P(X = 6) = \binom{10}{6} 0.5^6 (1 - 0.5)^{(10-6)} \tag{13}$$
$$= \frac{10!}{6!4!} 0.5^6 0.5^4 \tag{14}$$
$$= 210(0.5)^1 0 \tag{15}$$
$$= 0.205 \tag{16}$$

Getting 3 correct answers in a multiple choice 5 question exam (the student is unprepared. Each question has 5 possibilities (a,b,c,d,e)). Probability of getting a correct answer at any question is 1/5. P(3 correct answers in a 5 question exam)

$$P(X = 6) = \binom{5}{3} \left(\frac{1}{5}\right)^3 (1 - \frac{1}{5})^{(5-3)} \tag{17}$$
$$= \frac{5!}{3!2!} 0.2^3 0.8^2 \tag{18}$$
$$= 10(0.2)^3 (0.8)^2 \tag{19}$$
$$= 0.0512 \tag{20}$$

Probability of getting at least 3 correct answers out of five. This equivalent to find the P(3 correct answers) + P(4 correct answers) + P(5 correct answers) = $P(X = 3) + P(X = 4) + P(X = 5) = 0.051 + 0.006 + 0 = 0.057$

You can find the mean, variance, standard deviation, maximum usual value and minimum usual value for the binomial distribution with special formulas

Mean, Expected Value

$$E = \mu = \sum XP(X) = np$$

Variance

$$\sigma^2 = \sum[(X^2P(X))] - \mu^2 = npq$$

Standard Deviation

$$\sigma = \sqrt{\sum[(X^2P(X))] - \mu^2} = \sqrt{npq}$$

Outliers

1. Maximum Value: $np + 3\sqrt{npq}$
2. Minimum Value: $np - 3\sqrt{npq}$

Example 7: A study shows that 10% of Americans adults are left-handed. A statistics discussion has 25 students in attendance. What is the probability 3 people are left-handed.
Part 1. P(3 people are left-handed)
Information:

- X is the number of left-handed people in class
- $n = 25$
- $X = 3$
- $p = 0.1$
- $q = 0.9$

$$\begin{aligned} P(X = 3) &= \binom{25}{3}\left(\frac{1}{10}\right)^3(1 - \frac{9}{10})^{(25-3)} && (21) \\ &= \frac{25!}{3!22!}0.1^3 0.9^{22} && (22) \\ &= 10(0.2)^3(0.8)^2 && (23) \\ &= 0.226(Rounded) && (24) \end{aligned}$$

Part 2. Find the mean and standard deviation of left handed students in the discussion.

1. $\mu = np = 25(0.1) = 2.5$ left handed students
2. $\sigma = \sqrt{npq} = \sqrt{25(0.1)(0.9)} = 1.5$ left handed students

Part 3. Would it be unusual to find a discussion of 25 students with 5 left-handed students?

1. Maximum Value: $np + 3\sqrt{npq} = 2.5 + 3(1.5) = 7$
2. Minimum Value: $np - 3\sqrt{npq} = 2.5 - 3(1.5) = -2$

5 is an usual value because it is between the max and min.

5.4 The Poisson Distribution.

Description of the Poisson Distribution

- Discrete probability distribution.
- The random variable is the number of occurrences (counts) of an event in an interval
- The interval can be: time, distance, area, volume, or some similar unit.

EXAMPLES:

- Number of earthquakes (at least 6.0 on the Richter scale) in the last 100 years
- Number of patients arriving at the Emergency Room on Fridays between 10:00 pm and 11:00 pm
- Number of buses that pass a bus stop within an hour

Poisson Distribution: Requirements

- Random variable X is the number of occurrences of an event over some interval
- The occurrences must be random
- The occurrences must be independent

To find probabilities we must use the Poisson probability distribution, which can be seen as

$$P(X = x) = \frac{\mu^x \exp{-\mu}}{x!} \tag{25}$$

where $x = 0, 1, 2, 3, 4, ...$ and $e \approx 2.71828$ (Euler's number) The Poisson distribution only depends on μ (the mean of the process).

You can find the mean, variance, standard deviation, maximum usual value and minimum usual value for the Poisson distribution with special formulas

Mean, Expected Value

$$E = \mu = \sum XP(X) = \mu(\text{\# occurrences within interval})$$

Variance

$$\sigma^2 = \sum[(X^2P(X))] - \mu^2 = \mu(\text{Variance is equal to the Mean})$$

Standard Deviation

$$\sigma = \sqrt{\sum[(X^2P(X))] - \mu^2} = \sqrt{\mu}((\text{Standard deviation is the square root of the mean})$$

EXAMPLE Beginning next class