Chapter 5 Lecture Notes

Matt Warner

5.2 - Probability Distributions for Descrete Random Variables

Chapter 4

- Samples space = S = the set of all possible outcomes from an experiment
- Events (Written as A,B,C,etc.) were subsets of outcomes from a sample space
- NEw events were formed by using the operations $\bigcup$ (union), $\bigcap$ (intersection).
  and $^{\complement}$ (complement).
- P(A) = the probabilty that some outcome (or any outcome) in the event would occur

Now

- We will express our work using random variables
- **Random variable** (written as X) is a variable whose possible values are determined by random chance. (The possible values of X are numerical outcomes of a random phenomenon)
- 2 types of random variables
  - **Discrete** - values of X are typically whole numbers such as 0,1,2,3,etc.
    (usually associated with counting, e.g., "the number of")
  - **Continuous** - values of X are typically every value from an interval on the number line
    (usually associated with measuring, e.g., heights, weights, times, etc.)

Question 1

let the discrete random variable X = the roll of an unbalanced or unfair or rigged die, suppose the probability distribution of X is given by the following table.

x	1	2	3	4	5	6
$p(x)$	0.40	0.20	0.12	0.10	0.14	?

Note:-

- X would be **discrete** because its possible values are the whole numbers 1, 2, 3, 4, 5, 6
- Events are written as $X = 2, X < 4, 2 \leq X, X < 6, etc.$

Notation

p(x) is shorthand for P(X = x)

e.g., p(2) = P(X = 2)

Find p(6)

That is,

$$P(X = 6)$$

Sum all the other values in the table

$$= .40 + .20 + .12 + .10 + .14 = .96$$

So,

$$p(X = 6) = 1 - 0.96$$
$$= .04$$

Find $\mathbf{P(1 < X \leq 4)}$

That is,

$$P(2) + P(3) + P(4)$$

So,

$$P(1 < X \leq 4) = .20 + .12 + .10$$

$$= 0.42$$

Find the probability that the roll is at least 4

That is,

$$P(X \geq 4)$$

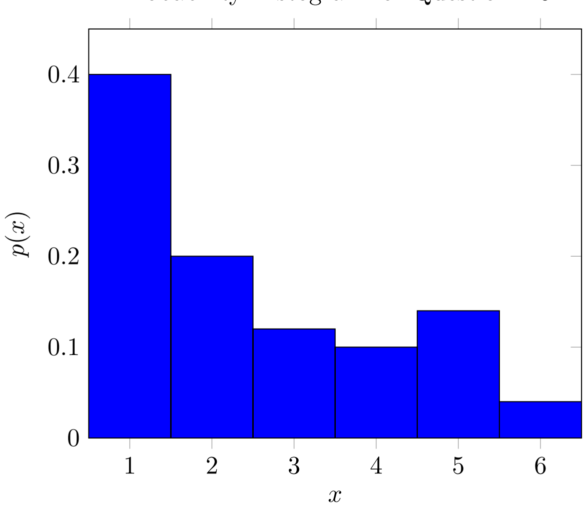
So,

$$P(X \geq 4) = p(4) + p(5) + p(6)$$

$$= .10 + .14 + .04$$

$$= 0.28$$

Probability Histogram for Question 1.0



Section 5.3 - Mean, Variance, and Standard Deviation for a Discrete R.V.

Important summaries of a data set are

- Sample mean = $\bar{x}$ (This indicates the center of the **data**)
- Sample standard deviation = s (This indicates the amount of spread among the **data**)

Likewise, a random variable or a probability distribution has similar summaries

Definition

Let X be a discrete random variable with probability mass function p(x)

- The **mean value** (μ) or the **expected value** (E(X)) of X is

Symbol Calculation

$$\mu \text{ or } E(X) = \sum_{all\ x} [x \cdot p(x)]$$

- The **variance** of X is

Symbol Calculation

$$\sigma^2 \text{ or } Var(X) = \begin{cases} \sum_{all\ x} \left[(x-\mu)^2 \cdot p(x)\right] & \text{(Definition)} \\ E\left(X^2\right) - \mu^2 & \text{(Short Cut)} \end{cases}$$

- The **standard deviation of X is**

$$\sigma = \sqrt{\sigma^2}$$

Note:-

- The mean, μ, gives the value of X that we would expect to see, on average.
  This indicates the center of a probability histogram
- The standard deviation, σ, measures the amount of spread among the values of X or the amount of spread exhibited by the probability histogram.
- If σ is
  - Large, then the values of X and/or the probability histogram has more spread
  - Small, then the values of X and/or the probability histogram has less spread.

Question 2

Let the discrete random variable X = the roll of an unbalanced or unfair or rigged die. Suppose the probability distribution of X is given by the following table

x	1	2	3	4	5	6
$p(x)$	0.40	0.20	0.12	0.10	0.14	0.04

Find the mean value of X (also called the expected value of X)

We can use the formula

$$\sum [x \cdot p(x)]$$

So,

$$\mu = 1(.40) + 2(20) + 3(.12) + 4(.10) + 5(.14) + 6(.04)$$

$$= 2.50$$

Find $E(X^2) = \sum \left[x^2 \cdot p(x)\right]$

That is,

$$1^2(.40) + 2^2(.20) + 3^2(.12) + 4^2(.10) + 5(.14) + 6^2(.04)$$

$$= 8.82$$

Find the variance of X using the short cut

$$Var(X) = E(X^2) - \mu^2$$

So,

$$Var(X) = 8.82 - 2.50^2 = 2.57$$

Find the standard deviation of X

Using the formula

$$S = \sqrt{S^2}$$

We get

$$\sqrt{2.57} = 1.60$$

Section 5.4 - The Binomial Distribution

Binomial Experiment

- Any experiment or situation that satisfies the following

- There is a known number of trials (denoted by n)
- Each trial results in a success or a failure
- P(success) is the same for every trial (denoted by p)
- The trials are independent. That is, the outcome of one trial wil not influence or affect the outcome of another trial

Note:-

The values of n and p are called the **parameters** of the Binomial experiment

Binomial Random Variable

- X = the total number of success oberserved during a Binomial experiement
- Possible values of X are 0, 1, 2, 3, . . ., n

Question 3

For each of the following decide whether or not the random variable is a Binomial random variable

1. A particular variety of seed has an 80% germination rate.
   Let X = the number of seeds out of ten that germinate

   - n = number of seeds (10),

   - each seeds has either a success or failure (germinating or not)

   - p = .80 (probability of success (germinating). This is also the same p value for every trial)

   - seeds are independent

   So yes, This is a Binomial random variable

2. Five cards are drawn without replacement from a deck.
   Let X = the number of red cards drawn

   - n = 5 cards

   - there does exist either a success or failure

   - p is not the same everytime since the cards are not replaced.

   - this also means that the cards are not independent

   Therefore, this is not a binomial random variable.

3. You take a multiple choice quiz with five questions, each containing choices (a) to (d), by guessing
   Let X = the number of correct answers

   - n = 5

   - each trial does have a success or failure

   Therefore, this is not a binomial random variable

Binomial Probability Mass Function

- $X \sim Bin(n, p)$ means that X is a Binomial random variable with parameters n and p
- $p(x)$ = the probability of obtaining exactly x successes among the n trials

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

Where

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

$$n! = n \cdot (n-1) \cdot (n-2) \ldots 2 \cdot 1$$

Question 4

Suppose that you take a multiple choce quiz with $n = 5$ questions, each containing choices (a) to (d), by guessing (so that $p = 0.25$). Let X = the number of correct answers. Here we know that $X \sim Bin(n = 5, p = 0.25)$

Find the probability of getting exactly one correct answer.

That is,

$$P(X = 1)$$

Using the formula

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

We get

$$\binom{5}{1}(.25)(1-.25)^{5-1} = \frac{5!}{1!(5-1)} \cdot 0.25^1 \cdot 0.75^4$$

$$P(X = 1) = .3955$$

Find the probability of getting exactly two correct answers

That is

$$P(x = 2)$$

So,

$$\binom{5}{2}.25^2(1-.25)^{5-2} = \frac{5!}{2!(5-2)!} \cdot 0.25^2 \cdot 0.75^3 = 10 \cdot 0.25^2 \cdot 0.75^3 = 0.2637$$

Find the probability of getting at most one correct answer

That is

$$P(X \leq 1)$$

This equates to

$$P(x = 0) + P(x = 1)$$

So,

$$\binom{5}{0} 0.25^0 (1 - .25)^{5-0} + \binom{5}{1} (0.25)(0.75)^4 = P(x \leq 1)$$

$$.2373 + .3955 = P(X \leq 1)$$

$$P(X \leq 1) = 0.6328$$

Find the probability of getting at least four correct answers

That is

$$P(x \geq 4)$$
$$= P(x = 4) + P(x = 5)$$

So,

$$P(x = 4) = \frac{5!}{4!1!} 0.25^4 \cdot .75^1$$

$$P(x = 5) = \frac{5!}{5!0!} 0.25^5 \cdot 0.75^0$$

$$= .0146 + .0010 = .0156$$

Question 5

Suppose that $\mathbf{X \sim Bin(15, 0.40)}$. Use the tables of Binomial Distribution Cumulative Probabilities to find each of the following

$$n = 15, \qquad p = 0.40$$

a) $P(X \leq 7) = 0.787$

b) $P(X > 5) =$ using the table, we see that P(X $\leq$ 15) = 1, and P(X< 5 = .403) so, P(X > 5 = .597)

c) P(X $\leq$) = .027

d) P(2 < $x \leq$ 6) = P(X $\leq$ 6) - P($X \leq 2$) = .610 - .027 = .583

e) P(5 $\leq X \leq$ 9) = .966 - .217 = .749

f) P(3 < X < 7) = P(X <7) - P($X < 3$) = .610 - .091 = .519

g) P($X = 4$) .217 - .091 = .126

Binomial Mean, Variance, & Standard Deviation

- If X $\sim Bin(n, p)$ then,

- Mean:
$$\mu = n \cdot p$$
- Variance:
$$\sigma^2 = n \cdot p \cdot (1 - p)$$
- Standard Deviation:
$$\sigma = \sqrt{n \cdot p \cdot (1 - p)}$$

Question 6

Suppose that you take a multiple choice quiz with n

A) **How many questions would you expect to get correct, i.e., what is the mean value of X**

Using the formula

$$\mu = np$$

We get

$$\begin{aligned} n \cdot p &= 5(.25) \\ &= 1.25 \end{aligned}$$

Find the standard deviation of the number of correct answers

First we find the variance

$$\begin{aligned} \sigma^2 &= np(1 - p) \\ &= 5(.25)(.75) \\ &= .9375 \end{aligned}$$

Now we can find the Standard Deviation

$$\begin{aligned} \sigma &= \sqrt{.9375} \\ &= .9682 \end{aligned}$$

Summary - Using the Binomial Tables

Probability mass function for the Binomial distribution

$$p(x) = \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x} \qquad \text{for x = 0,1,2,\ldots,n}$$

The probability mass function is used to calculate the probability of exactly a successes,

$$P(X = a)$$

However, the Binomial tables are cummulative tables that provide the probability of at most a successes,

$$P(X \leq a)$$

Below is a summary of how to use the tables to obtain various probabilities. You should study these until you understand them . They are not on the formula sheet and won't be provided at the exams.

- $P(X \leq a) =$ table entry for a
- $P(X < a) = P(X \leq a-1)$ e.g. $P(X < 5) = P(X \leq 4)$
- $P(X \geq a) = 1 - P(X \leq a-1)$ e.g. $P(X \geq 5) = 1 - P(X \leq 4)$
- $P(x > a) = 1 - P(X \leq a)$ e.g. $P(X > 5) = 1 - P(X \leq 5)$
- $P(X = a) = P(X \leq a) - P(X \leq a-1)$ e.g. $P(X = 5) = P(X \leq 5) - P(X \leq 4)$
- $P(X \neq a) = 1 - P(X = a)$
- $P(a < X \leq b) = P(X \leq b) - P(X \leq a)$ e.g. $P(2 < X \leq 5) = P(x \leq 5) - P(X \leq 2)$
- $P(a \leq X \leq b) = P(X \leq b) - P(X \leq a-1)$ e.g. $P(2 \leq X \leq 5) = P(X \leq 5) - P(X \leq 1)$
- $P(a \leq X < b) = P(X \leq b-1) - P(X \leq a-1)$ e.g. $P(2 \leq X < 5) = P(X \leq 4) - P(X \leq 1)$
- $P(a < X < b) = P(X \leq b-1) - P(X \leq a)$ e.g. $P(2 < X < 5) = P(X \leq 4) - P(X \leq 2)$