Stat 201: Statistics I
Chapter 5

Chapter 5
Discrete Probability Distributions

Section 5.1
Probability Distributions

Random variables

A **random variable** is a variable that has a numeric value determined by chance from a range of possible values.

- An outcome of a trial
- Usually designated with a capital letter (X, Y, etc.)
- Lowercase letters refer to specific values of the random variable
- Thus, $P(X = x)$ means the probability that the random variable X takes the specific value x.

Example

- $X =$ the number of heads from three coin flips
- $Y =$ the sum of two dice
- $Z =$ the midterm score of a randomly selected student
  - A students grade (A, A-, B+, etc.) can not be used as a random variable because it is not numeric,
  - ... Unless, the grade is coded as a number (i.e. A $=$ 4.0, A- $=$ 3.7, etc.)

Types of random variables

Recall, numeric variables can be classified as **discrete** or **continuous**. Random variables also can be either discrete or continuous.

Example

Discrete random variables:

- Number of heads on three coin flips
- Number of defective insulin test strips in a box of 50
- Number of customers to enter a store in the next 10 minutes

Continuous random variables:

- Height or weight of a test subject
- Survival time of a cancer patient
- Price of a company's stock at a particular moment

Probability distributions

The collection of probabilities of all the possible values of a random variable is known as the **probability distribution** of the random variable.

- Each probability is between 0 and 1
- The probabilities must add up to 1
- Often displayed in tables (if practical)

Probability distributions, example

Example

A restaurant wants to track its taco sales. It records how many tacos customers order with each visit. The results are in the table,

Number of tacos	Probability
0	0.35
1	0.2
2	0.3
3	0.1
4	0.05

Is this a probability distribution?

- These are probabilities of every possible outcome of a trial (customer making an order).
- The probabilities add to 1.
- It is a probability distribution.

Probability distributions, example

Example

The restaurant introduces a new "Super Taco" (beef, chicken and shrimp). It wonders if larger groups are more likely to order the new item. The results are in the table,

Number in group	Probability of ordering
1	0.05
2	0.03
3	0.04
4 or more	0.15

Is this a probability distribution?

- The probabilities are for a different event (ordering a Super Taco) than the values (number in group).
- The probabilities do not add to 1.
- It is not a probability distribution.

Event probabilities

To calculate the probability of an event given a probability distribution, simply add the probabilities of the outcomes which comprise the event.

Example

Number of tacos	Probability
0	0.35
1	0.2
2	0.3
3	0.1
4	0.05

What is the probability of a customer ordering less than two tacos?

$$P(X < 2) = P(X = 0 \text{ or } 1) = P(0) + P(1) = 0.35 + 0.2 = 0.55$$

Weighted means

A **weighted mean** is the mean of values that are not considered equally, or do not have equal importance.

- Each value has an associated weight, which is its relative importance.
- To calculate, let x_i be a value and w_i its weight.

$$\mu_w = \frac{\sum w_i \times x_i}{\sum w_i}$$

Example

Fiona buys 4 lbs. of hamburger at \$4.89 / lb. and 2 lbs. of steak at \$11.99 / lb. What is the average price per pound she is paying?

$$\$/lb. = \frac{4 \times 4.89 + 2 \times 11.99}{6} = \frac{43.54}{6} = 7.26$$

Mean of probability distributions

The mean of a probability distribution is a weighted mean of the possible values, with the probability of each as its weight.

- Since the sum of probabilities of a distribution is always 1, the divisor of the weighted mean is 1 which we can ignore.
- Thus, the mean is

$$\mu = \sum x_i \cdot P(x_i)$$

The mean of a probability distribution is also known as the **expected value** of the random variable.

- Denoted with an "E", as in

$$\mathbb{E}(X) = \mu$$

Mean, example

Example

Number of tacos	Probability
0	0.35
1	0.2
2	0.3
3	0.1
4	0.05

What is the mean number of tacos ordered at the restaurant? That is, how many tacos should the restaurant expect each customer to order?

$$\begin{aligned}\mathbb{E}(X) = \mu = & \sum x_i \cdot P(x_i) \\ = & \; 0 \cdot 0.35 + 1 \cdot 0.2 + 2 \cdot 0.3 + 3 \cdot 0.1 + 4 \cdot 0.05 \\ = & \; 0 + 0.2 + 0.6 + 0.3 + 0.2 \\ = & \; 1.3\end{aligned}$$

Standard deviation of probability distributions

Similarly, variance of a probability distribution is the weighted mean of difference from the mean squared and standard deviation is the square root of variance.

Thus,

$$\sigma^2 = \sum (x_i - \bar{x})^2 \cdot P(x_i)$$

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

Standard deviation, example

Example

What is the standard deviation of number of tacos ordered at the restaurant?

$$\begin{aligned} \sigma^2 = &\quad \sum (x_i - \bar{x})^2 \cdot P(x_i) \\ = &\quad (0 - 1.3)^2 \cdot 0.35 + \cdots + (4 - 1.3)^2 \cdot 0.05 \\ = &\quad 0.5915 + \cdots + 0.3645 \\ = &\quad 1.41 \\ \sigma = \sqrt{\sigma^2} = &\quad 1.19 \end{aligned}$$

Unusual events rule of thumb

Recall, the range rule of thumb for unusual values are those values more than 2 standard deviations away from the mean. In other words, x is unusual if

$$x < \mu - 2\sigma \qquad \text{or} \qquad x > \mu + 2\sigma$$

Example

From the probability distribution of tacos ordered, $\mu = 1.3$ and $\sigma = 1.19$. What would be an unusual amount of tacos to order?

- The lower bound for unusual values is $\mu - 2\sigma = -1.08$. Since you can't order negative tacos, there is not an unusually low number of tacos to order.
- The upper bound for unusual values if $\mu + 2\sigma = 3.86$. Thus, 4 (or more) tacos is an unusually high number of tacos to order.

More precise definition for unusual events

If the probability of a random variable being equal to an event or takes a value more extreme than than the event is less than some threshold, usually 0.05, then the event is an **unusual event**.

That is, if x is an extreme value if

$$P(X \leq x \text{ or } X \geq x) < 0.05$$

$$P(X \leq x) + P(X \geq x) < 0.05$$

$$P(X \leq x) < 0.025 \qquad \text{or} \qquad P(X \geq x) < 0.025$$

Unusual events, example

Example

Tom and Barbara collect eggs from their chickens everyday. The number of eggs they collect follows the following distribution,

Eggs (x)	0	1	2	3	4	5	6	7	8	9
$P(x)$	0.01	0.03	0.1	0.2	0.3	0.2	0.1	0.04	0.02	0

Is collecting no eggs unusual? Is collecting 7 eggs unusual?

- $P(X \leq 0) = P(0) = 0.01 < 0.025$
- $P(X \geq 7) = P(7) + P(8) + P(9) = 0.06 \not< 0.025$

Rare event rule

The **rare event rule** says that if observed results are unusual, given an assumed probability distribution, then perhaps the assumption is wrong.

Example

Recall the example of flipping a coin 1000 times. Under the assumption the coin is fair ($P(H) = P(T) = 1/2$), the expected number of heads is 500.

- Getting 523 heads is not unusual ($P(X \geq 523) = 0.077$). There is no reason to think the coin is not fair.
- Getting 46 heads is unusual ($P(X \leq 46) = 6.23 \times 10^{-222}$). We would be justified in questioning the assumption that the coin is fair.

Section 5.2
Binomial Probability Distributions

Binomial distribution

A **binomial distribution** is a probability distribution representing the number of "successes" in a fixed number of trials with two possible outcomes.

- The term "success" is traditional, but can refer to any outcome of interest.

Example

- The number of students passing the midterm in class
  - Passing the midterm is a "success", failing it is a "failure"
- The number of heads in three coin flips
  - Getting a head is a "success", getting a tail is a "failure"
- The number of of car crashes that result in fatalities
  - A fatality is a "success", no fatalities is a "failure"

Requirements for binomial distributions

There are four requirements to be considered a binomial distribution:

- The distribution must represent a fixed number of trials
- Each trial must have exactly two possible outcomes (success / failure)
- Each trial must be independent of the other trials
- Each trial must have the same probability for success

The last two requirements are often summarized as "independent and identically distributed" and abbreviated as "iid".

Requirements for binomial distributions, example

Example

Recall the Metro State Statistics Club, with 6 male members and 4 females members. If the three officer positions are selected randomly, does the number of women selected follow a binomial distribution?

- No. The events are not independent.

Example

An instuctor is trying a new method of testing students. A test is given once a day for five days, with opportunities to practice and ask questions in between each test. Does the number of times a student passes the test follow a binomial distribution?

- No. The probability of success changes (hopefully) with each test.

Requirements for binomial distributions, example

Example

Recall the Youth Risk Behavior Survey (YRBS) which found the probability of a teenaged driver had texted or emailed while driving was 0.404. Suppose this is the probability for the population of teenaged drivers. Suppose 30 teenaged drivers are selected at random. Does the number of those drivers that had texted or emailed while driving follow a binomial distribution?

- Yes.
    - There is a fixed number of trials (30).
    - Each trial has only two possible outcomes (had or had not texted).
    - Each trial is independent.
    - Each trial has the same probability of "success."

Notation for binomial distributions

$$X \sim \text{Binom}\,(n, p)$$

- The random variable X follows a binomial distribution with n trials and p probability of a success.
- $q = 1 - p =$ the probability of failure
- $P(X = x)$ is the probability of getting exactly x successes. x must be between 0 and n.

Example

The probability a teenaged driver had texted or emailed while driving is 0.404. If the random variable Y is the number of teenaged drivers who had texted or emailed while driving out of a sample of 30,

$$Y \sim \text{Binom}\,(30, 0.404)$$

Probabilities for binomial distributions

The formula for binomial probabilities is

$$P(X = x) = \binom{n}{x} p^x q^{n-x}$$

- $\binom{n}{x}$ is read as "n choose x". It is the number of ways to get x successes in n trials. For example, we previously determined that there were three ways to get two heads in three flips, $\{$ HHT, HTH, THH $\}$. Thus, $\binom{3}{2}$ is 3.
- $p^x q^{n-x}$ is the probability of getting x successes and $n - x$ failures in one particular order.

Practice: cancer screening

Recall the cancer screening example. Out of 1000 people screened, the probability of a positive test was 0.1. Suppose a sample 40 of those screened were randomly selected.

The random variable X is the number who tested positive in the sample. Does X follow a binomial distribution?

- Yes. All the requirements are met.

What is the notation for X?

- $X \sim \text{Binom}\,(40, 0.1)$

What is the probability that exactly 5 people in the sample tested positive?

- $P(X = 5) = 0.165$

Parameters for Binomial Distributions

For $X \sim \text{Binom}\,(n, p)$,

- Mean:
$$\mathbb{E}(X) = \mu = np$$
- Variance:
$$\sigma^2 = npq$$
- Standard deviation:
$$\sigma = \sqrt{\sigma^2} = \sqrt{npq}$$

Parameters, example

Example

The random variable Y is the number of teenaged drivers who had texted or emailed while driving out of a sample of 30,

$$Y \sim \mathrm{Binom}\,(30, 0.404)$$

Parameters of Y:

- $\mathbb{E}(Y) = \mu = np = (30)(0.404) = 12.12$
- $\sigma^2 = npq = (30)(0.404)(1 - 0.404) = 7.22$
- $\sigma = \sqrt{\sigma^2} = \sqrt{7.22} = 2.69$

Unusual values

For a given binomial distribution, the boundaries for unusual values can be found. From the range rule of thumb, the lower boundary is $\mu - 2\sigma$ and the upper boundary is $\mu + 2\sigma$.

Example

For $Y \sim \mathrm{Binom}(30, 0.404)$, $\mu = 12.12$ and $\sigma = 2.69$. What are the unusual values?

- The lower bound for unusual values is $\mu - 2\sigma = 6.74$
- The upper bound for unusual values is $\mu + 2\sigma = 17.5$
- In a random sample of 30 teenaged drivers, it would be unusual to get 6 or fewer, or 18 or more, drivers who had texted or emailed while driving.

Section 5.3
Poisson Probability Distributions

Poisson Probability Distributions

A **Poisson distribution** represents the number of events occurring in a specified interval.

- Assumes events are random, independent and uniformly distributed.
- Intervals are using lengths of time, but other kind of intervals are possible.

Example

- The number of fish caught in the next hour
- The number of customers to come into a store between 1 and 6 pm
- The number of insects found in a square foot of grassland

Poisson notation

$$X \sim \mathrm{Pois}\left(\lambda\right)$$

- The random variable X has a Poisson distribution with an mean rate of events of λ (lambda).
- The mean rate (λ) is scalable. That is, if λ is the mean number of events per hour, $\mathrm{Pois}\left(\lambda/2\right)$ models the number of events in a half hour interval.

Poisson probability and parameters

The probability of x events in an interval,

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

- Mean:

$$\mathbb{E}(X) = \mu = \lambda$$

- Variance:

$$\sigma^2 = \lambda$$

- Standard deviation:

$$\sigma = \sqrt{\lambda}$$

Poisson, example

Example

Paul fishes for an hour every morning. Over the past month (30 days) he has caught 62 fish. Paul is having friends over for dinner tonight, so he needs to catch 4 fish.

What is the probability he catches 4 fish in an hour? Is that an unusually high number?

- Assuming the chances of catching a fish are the same every day, the mean rate of fish caught is $\lambda = \frac{62}{30} = 2.07$
- Probability of catching exactly 4 fish: $P(X = 4) = 0.097$
- Probability of catching four or more fish: $P(X \geq 4) = 0.156$
- Rule of thumb boundary: $\mu + 2\sigma = 2.07 + 2\sqrt{2.07} = 4.95$