Packet 2: Discrete Probability Distributions

Student Learning Outcomes:

- Determine if a proposed distribution is a valid probability mass function.
- Find the cumulative distribution function for a discrete random variable.
- Find the mean and variance of discrete random variables and their transformations.
- Be able to identify and use the binomial distribution.
- Be able to identify and use the geometric distribution.
- Be able to identify and use the hypergeometric distribution.
- Be able to identify and use the negative binomial distribution.
- Be able to identify and use the Poisson distribution.
- Find the mean and variance of the named discrete distributions.
- Use moment generating functions to find the moments of discrete distributions.

Types of variables

In general, data can be:

-
-

Example: People rate the statement "The Patriots will win the Super Bowl." using the following scale: Strongly Agree, Agree, Disagree, Strongly Disagree.

Example: A consumer agency measures the amount of time before a light bulb burns out.

To statistically analyze data, we need to pick a numerical summary of interest, so we need a way to assign a numeric value to each possible outcome.

____: A function that associates each element of the sample space with a number.

Notation:

Example (Opinion Ratings): S = {Strongly Agree, Agree, Disagree, Strongly Disagree}

Random variables are a function with an input that is an element of the sample space and an output that is a real number.

A random variable, X is ____________________ if the possible values of X are finite or countable.

Example:

Example:

Example:

Probability Distributions for Discrete Random Variables

____ of a discrete random variable describes how likely the possible values of x (i.e. possible inputs) are to occur.

Notation:

Example: Let X = the number of correct answers on an exam with 20 questions.

$$p(x) = P(X = x) = \begin{cases} .1 & x = 20 \\ .2 & x = 19 \\ .5 & x = 18 \\ .1 & x = 17 \\ .05 & x = 16 \\ .05 & x = 15 \\ 0 & x < 15 \end{cases}$$

Properties of Probability Mass Functions:

1.

2.

Anything that satisfies these criteria is a valid probability mass function.

Example: Let X = the distance a golf ball is from the flag after your first swing, rounded to the nearest foot. Then $X \in 0, 1, 2, ..$ Suppose the pmf is given by:

$$P(X = x) = \begin{cases} 1/2 & x = 0 \\ 1/4 & x = 1 \\ 1/8 & x = 2 \\ 1/16 & x = 3 \\ ... & \\ (1/2)^{i+1} & x = i \end{cases}$$

Is this a valid pmf? Explain.

What is the probability that X is smaller than or equal to 2?

What is the probability that X is greater than 2?

Families of Distributions

Some probability mass functions are indexed by ____________________________, which are variables that help to define the pmf and that can take any one of several possible values. Each possible value of a parameter defines a different pmf.

Example:

By varying the parameter, we can create many different probability distributions. The set of all probability distributions obtained by varying a parameter is called a __________________ of distributions.

Example:

Notation:

Another way to describe Bernoulli random variables is as ________________________________. An indicator variable random variable (sometimes denoted I_A) takes on the value 1 if the experiment is considered a success, and it takes on the value 0 if the experiment is considered a failure.

Example: A multiple choice test has 20 questions, each worth one point. Let:

$$X_j = \begin{cases} 1 & \text{if a student is correct on question j} \\ 0 & \text{otherwise} \end{cases}$$

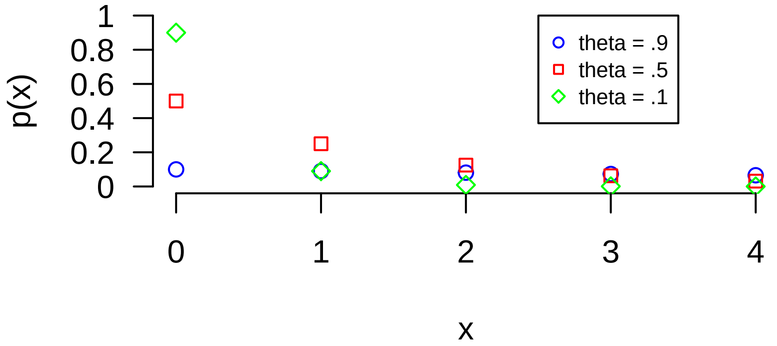
and let Y be a student's total score ($Y = X_1 + X_2 + . + X_{20}$). Assume that X_j ~ **Bernoulli(θ)** are independent. What is the probability that a student gets a perfect score?

What if $\theta = 0.8$?

What if $\theta = 0.5$?

Graphs of discrete distributions

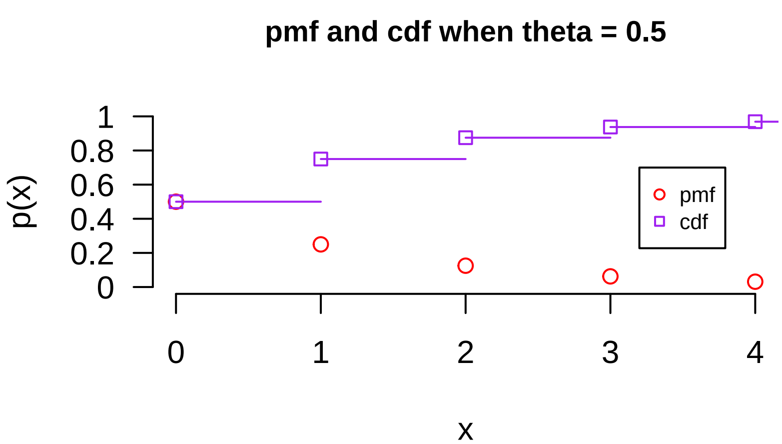
Example: $p(x) = (1-\theta)\theta^x$, θ in $(0,1)$. where x $\in \{0, 1, 2, \ldots\}$



____ of a discrete random variable X, provides the probability that the observed value of X will be at most x.

Notation:

Example:



Expectation and Variance

Expected Value

The ________________________ of a random variable is the long term average for the random variable. If X is a discrete random variable with set of possible values D and pmf p(x), the expected value of X is found by:

Example:

$$p(x) = P(X = x) = \begin{cases} .1 & x = 20 \\ .2 & x = 19 \\ .5 & x = 18 \\ .1 & x = 17 \\ .05 & x = 16 \\ .05 & x = 15 \\ 0 & x < 15 \end{cases}$$

Example: Let X be an indicator random variable with $p(x) = \theta$. What is $E(X)$?

Example: Let $p(x) = (1-\theta)\theta^x$ for $\theta \in (0,1)$ and x = $\{0,\ 1,\ 2,\ \ldots\}$. Find $E(X)$.

Let X be a discrete random variable with a set of possible values D and pmf p(x). Also let h(X) be some function of our random variable X. The expected value of h(X) is:

Example:

$$X = \begin{cases} 1 & \text{with probability } \theta \\ 0 & \text{with probability } 1-\theta \end{cases}$$

What is the expected value of $h(X) = X^2$?

In general, for any numbers a and b and random variable X:

Variance is a measure of the spread for our random variable.

Let X be a discrete random variable with set of possible values D, pmf p(x), and mean μ. The variance of X can be found by:

Proposition: $Var(a + bX) = b^2 * Var(X)$

____ is the square root of the variance.

Both variance and standard deviation are measures of the spread of a distribution. They represent the "expected squared deviation from the mean" or, roughly speaking, the average distance of the observations from the mean.

Example: Let X ~ Bernoulli(θ). The pmf of X can be written as:

$p(x) = \theta^x(1-\theta)^{1-x}$, X = {0, 1}. Find E(X), Var(X) and SD(X).

Example: Find the mean and variance for the random variable X (see table). Use your result to find the mean and variance of the the random variables

Y_1 = 3 + X, Y_2 = 2X and Y_3 = 3 + 2X.

X	1	2	3
P(X=x)	0.2	0.6	0.2

Named Discrete Distributions

The Binomial Distribution

Recall our previous example(s) about test scores:

Example: A multiple choice test has 20 questions, each worth one point. Let:

$$X_j = \begin{cases} 1 & \text{if a student is correct on question j} \\ 0 & \text{otherwise} \end{cases}$$

and let Y be a student's total score ($Y = X_1 + X_2 + . + X_{20}$). Assume that $X_j \sim$ **Bernoulli(.8)** are independent.

What is the probability that a student gets a perfect score?

For this student, what is the probability of getting 18 questions correct?

- For each term, ________ questions are answered correctly and ________ are not. Since the events are independent, each term will equal:
- How many terms are there? In other words, how many ways are there to answer ________ questions correctly and ________ incorrectly? This question can be answered by considering the number of ____________________________________.

The number of ways to answer 18 out of 20 questions correctly is:

The probability of getting 18 out 20 questions correct is:

In this example what was of interest is the probability for a specific __________ of correct questions. This will be the case with many of our named discrete random variables.

Recall the pmf for the Bernoulli family of distributions:

$$P(X = x) = \begin{cases} p, & x = 1 \\ 1 - p, & x = 0 \end{cases}, \text{ for } p \in (0, 1)$$

Often the random variable X is defined to equal 1 if some condition is met, or in other words the experiment yields a _____________, and X is equal to 0 otherwise (the experiment yields a _____________). Note: success does not have to be a 'good' thing. It might be something bad, but if we find what we are looking for, we call it a success.

Now imagine that we conduct the same experiment (with success probability p), n times, with each of the n trials being independent of the other trials. Then...

The probability distribution for a **Binomial(n, p)** random variable is given by:

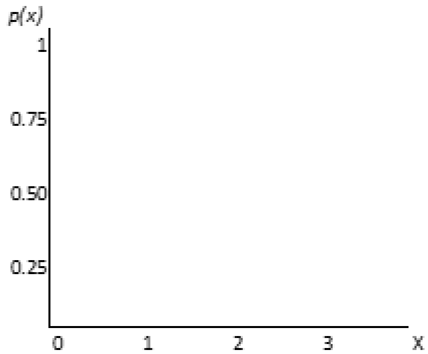
Notes:

1. If n=1, the Binomial distribution is the same as the Bernoulli distribution.

   That is: Binomial(1, p) = Bernoulli(p).

2. Suppose X_i are independent $Bernoulli(p), i = 1...n$, then $Y = \sum_{i=1}^{n}(X_i) \sim Binomial(n, p)$

Example: X~Binomial(3,0.1). What is P(X=x)?



.

What is the probability of at least two successes?

What is the probability of at least one success?

Any variable that counts the number of successes from a binomial experiment is a binomial random variable.

An experiment is a binomial experiment if it meets each of the following conditions:

1. The number of trials (denoted n) is:
2. There are 2 possible outcomes of each trial:
3. Outcomes are independent from one trial to the next.
4. The probability of "success" (denoted p) remains the same from one trial to the next.

Example: A ten-question quiz has five True/False questions and 5 multiple-choice questions, each with four possible choices. Suppose a student randomly picks an answer for each of the 10 questions. Let X = {the number of correct answers}. Is X a binomial random variable? If so, describe its distribution. If not, say why not.

Mean and Variance of the Binomial Distribution

If X follows a Binomial distribution with n independent trials and probability of success p for each trial, then X has:

Mean:

Variance:

Derivation of E(X):

Theorem: Let $Y_1, Y_2, ..., Y_n$ be random variables with $E(Y_i) = \mu_i$, then

$$E[\sum_{i=1}^{n} Y_i] = \sum_{i=1}^{n} E[Y_i] = \sum_{i=1}^{n} \mu_i$$

If $Y_1, Y_2, ..., Y_n$ are independent, then

$$Var[\sum_{i=1}^{n} Y_i] = \sum_{i=1}^{n} Var[Y_i]$$

In addition to the derivation of E(X) we just found, we could also use the fact that a binomial random variable is the sum of n independent Bernoulli random variables to derive this expectation. Recall that $E[Bernoulli(p)] = p$. Let $Y = \sum_{i=1}^{n}(X_i)$ where X_i are iid Bernoulli(p). Derive E(Y).

We can use this same method to derive the variance of a binomial random variable. Recall that $Var[Bernoulli(p)] = p(1-p)$

Example: Suppose that about 10% of Americans are left-handed. Let X = the number of left-handed Americans in a random sample of 12 Americans.

a. What is the distribution of X?

b. What are the mean and variance ?

c. What is the probability that the sample contains 2 or fewer left-handed Americans?

d. Suppose we define a random variable Y = 12 - X. Describe in words what the random variable Y is measuring. What is it's distribution?

Hypergeometric Distribution

Example: We are interested in learning about a student-run club on campus, which has 20 members. In this club, 10 members are women and 10 members are men. Suppose we sample one member at random and record

$$Y = \begin{cases} 1 & \text{if Female} \\ 0 & \text{if Male} \end{cases}$$

What distribution does Y follow?

Now suppose we sample 4 members at random *without replacement* and record X = {number of females}. What distribution does X follow?

What if the club had 20,000 members, half of which were females?

Note:

- We can use the Binomial Distribution
  1. If the population is at least 10 times as large as the sample.
  2. Or if we are sampling with replacement.

- When sampling without replacement from a small population, we cannot use the Binomial Distribution, instead we must use the ________________________ distribution.

Conditions for a Hypergeometric Distribution:

1. The population is finite and

2. There are 2 possible outcomes for each individual/object: success and failure:

3. A sample of n individuals are

If these conditions are met, then X = {number of success} is a hypergeometric random variable.

The distribution of X depends on three parameters:

Notation:

The probability distribution for a **Hypergeometric(n,r,N)** random variable is given by:

Mean:

Variance:

Notes:

- _______________ is the proportion of success in the population, meaning the probability of success. If we replace this term with ________ in the above formulas, then we get:
- This is almost the same mean and variance as the binomial distribution. The additional term in the variance is the ____.
- This factor is less than 1, so the hypergeometric distribution has a __________________________ than the binomial distribution.
- When n is a relatively small fraction of N, this factor is approximately 1, so the hypergeometric distribution would be essentially equivalent to the binomial distribution.

Example: The biology club weekend outing has two groups. One with 7 people will camp at Diamond Lake. The other group with 10 people will camp at Arapahoe Pass. Seventeen duffels were prepared by the outing committee, but 6 of these had the tents accidentally left out of the duffel. The group going to Diamond Lake picked up their duffels at random from the collection and started off on the trail. The group going to Arapahoe Pass used the remaining duffels. Let X = {number of Diamond Lake campers without tents}.

a. What distribution does X follow? Include the support of X.

b. What is the probability that at least 1 camper going to Diamond Lake will not have a tent in their duffel?

c. What is the probability that 6 campers going to Diamond Lake will not have a tent in their duffel?

d. What is the probability that all 7 campers going to Diamond Lake will not have a tent in their duffel?

e. What is the expected number of campers going to Diamond Lake that will not have a tent in their duffel?

f. What are the variance and standard deviation of X?

Geometric Distribution

For the binomial and hypergeometric distributions, we were trying to model a random number of successes in a specific number of trials. For the next two distributions, the geometric distribution and the negative binomial distribution, our goal will be to *model a specific number of successes when the number of trials is not fixed.*

> *Example*: Suppose that the probability of an engine malfunction during any one-hour period is p = 0.02. Note: "success" in this example is defined as the engine malfunctions. What is the probability that the engine survives until the 4th hour and then malfunctions?

Note that what this example shows us is that for each value of the random variable, X, we will have one success and $x - 1$ failures. When we reach our success criteria, our experiment ends. Thus it follows that

The probability distribution for a **Geometric(p)** random variable is given by:
Mean:
Variance:

Just as we have seen with our other distributions, there are certain conditions that need to be met to classify the probabilities of an experiment as having a geometric distribution.

Conditions for a Geometric Distribution: 1. Trials of the experiment are independent of each other. 2. There are two possible outcomes of each trial: success and failure. 3. The probability of success, p, remains the same from one trial to the next.

Derive the expected value of the geometric distribution.

Example: Suppose that 30% of the applicants for a certain industrial job possess advanced training in computer programming. Applicants are interviewed sequentially and are selected at random from the pool.

1. What is the probability that the first applicant with advanced training in programming is found on the third interview?

2. What is the probability that it will take more than 2 interviews to find the first applicant with advanced training?

3. What is the expected number of interviews that it will take before an applicant with advanced training in computer programming is found?

4. What is the variance for the number of interviews that it will take before an applicant with advanced training in computer programming is found?

Negative Binomial Distribution

Similar to the geometric distribution, is the negative binomial distribution. For these types of problems, instead of looking for a single success and stopping the experiment, we will now look for one *or more* successes before we stop the experiment. For example:

> Suppose that each year Indiana companies have a 0.66 probability of being awarded an army contract. What is the expected number of years it would take until 3 contracts were awarded to Indiana companies?

Conditions for a Negative Binomial Distribution:

1. The experiment consists of

2. There are 2 possible outcomes for each trail:

3. The probability of success

4. The experiment continues until a total of r successes have been observed.

If these conditions are met, then X = {number of trials needed for r successes to occur} is a **negative binomial** random variable with parameters: r (number of successes) and p (probability of success).

Notation:

Notes:

- In both the binomial and hypergeometric distributions, the number of trials is ____________________ and the number of successes is _________________.
- In the negative binomial distribution (like the geometric distribution), the number of trials is _______________ and the number of successes is _________________.
- Negative binomial distributions are also referred to as binomial waiting time distributions or Pascal distributions.

Example:

- Suppose we were interested in X = {the number of trials before the first success}. What is P(X=5)?
  - In this problem we are interested in ______________; meaning there were ______ failures followed by ______ success:

- Suppose we are interested in Y= {the number of trials before the second success} What is P(Y=4)?
  - In this problem we are interested in ______________; meaning there were ______ failures and ______ successes.
  - Note that, just like the geometric distribution, the last trial always has to be a ________________________. There are 3 possible sequences that would satisfy this outcome:

The probability distribution for a **NegBin(r, p)** random variable is given by:

Mean:

Variance:

Just as we saw with the relationship between the binomial distribution and the Bernoulli distribution, there is a special relationship between the geometric and negative binomial distributions.

Suppose X_i are iid $geometric(p), i = 1...r$, then $Y = \sum_{i=1}^{r}(X_i)$ is $NegBin(r, p)$

We can utilize this relationship to derive the expected value and variance of the negative binomial distribution.

Example: Blood type A occurs in about 41% of the population. A clinic needs 4 pints of type A blood. Each donor gives a pint of blood each time they donate. Let X be a random variable representing the number of donors needed to provide 4 pints of type A blood.

a. What is the distribution of X with support? What assumptions are you making?

b. Compute the $P(X = 4)$, $P(X = 5)$, $P(X = 6)$ and $P(X = 7)$.

c. What is the probability that the clinic will need between 4 and 7 donors to obtain the needed 4 pints of type A blood.

d. What is the probability that the clinic will need more than 7 donors to obtain the 4 pints of type A blood?

e. What are the expected value and standard deviation of X?

Poisson Distribution

The last discrete distribution we will discuss, called the *Poisson Distribution*, can be used to count the number of successes during a specified interval of time or space.

Examples:

The probability distribution for a **Poisson(λ)** random variable is given by:

Mean:

Variance:

Derive the mean of a Poisson(λ) random variable.

Example: At Burnt Mesa Pueblo, in one of the archaeological excavation sites, the artifact density (number of prehistoric artifacts per 10 liters of sediment) was 1.5. Let $X = 0, 1, 2, \ldots$ be a random variable that represents the number of prehistoric artifacts found in 10 liters of sediment.

a. Compute the probability that 0, 1 or 2 prehistoric artifacts will be found in 10 liters of sediment.

b. Compute the probability that 3 or more prehistoric artifacts will be found in 10 liters of sediment.

c. Let Y be the number of artifacts found in 50 liters of sediment. Find the expected value and variance for Y.

d. Compute the probability that 3 or more prehistoric artifacts will be found in 50 liters of sediment.

Moment Generating Functions

A moment generating function is a function that helps us calculate the moments of a distribution. Some examples of moments that we have already used this semester include:

Recall that these are useful for helping us find the Expected Value and the variance of a distribution. This means that moment generating functions will give us another way to find these values (and others) that may provide easier functions to work with summations and integrals.

To find a moment generating function $M_X(t)$*, we will need to find* $E[e^{tx}]$. Recall that to find the expected value of any function we need to multiply that function by the pmf and sum over all possible values of the random variable.

Example: Find the moment generating function for the binomial distribution.

The n^{th} moment of a distribution is defined as $E(X^n)$. To find the n^{th} moment using a moment generating function, you can take the n^{th} derivative of the moment generating function with respect to t, then set t = 0. For example to find find E(X), find $M_X'(0)$ or to find $E(X^2)$, find $M_X''(0)$.

Example: Use the Moment Generating Function on the previous page to find E(X) and $E(X^2)$ of the binomial distribution. Then find the Var(X). Do these match the values you were given when we discussed the binomial distribution?

Example: Let Y~Poisson(μ). Find the moment generating function of Y and use it to show that the mean and variance of this distribution are both μ.

Why is it that we can use this process with Moment Generating Functions to find moments? To answer this question, we need to recall the Taylor expansion of e^x.

This means that $M_X(t)$ can also be written as:

And when we take the derivative of the Moment Generating Function we get:

And finally when we evaluate the derivative of the Moment Generating Function at 0 we get: