Part 4: More of the Common Discrete Random Variable Distributions

Sections 3.6 & 3.7 Geometric, Negative Binomial, Hypergeometric

NOTE: The discrete Poisson distribution (Section 3.8) will be on midterm exam 2, not midterm exam 1.
The following common discrete random variable distributions will be on Midterm Exam 1:

- Discrete Uniform
- Binomial
- Geometric
- Negative Binomial
- Hypergeometric

The only other common discrete random variable we cover, the Poisson, will be part of Midterm Exam 2, not Midterm Exam 1.
Geometric Distribution

- Consider a sequence of independent Bernoulli trials with a success denoted by $S$ and failure denoted by $F$ with

$$P(S) = p \text{ and } P(F) = 1 - p.$$ 

Let $X =$ the number of trials until (and including) the first success.

Then, $X$ follows a Geometric Distribution.

The only parameter needed is the probability of a success $p$.

**NOTATION:**
If $X$ follows a geometric distribution with parameter $p$ (for probability of success), we can write

$$X \sim Geo(p) \text{ for } X = 1, 2, 3, \ldots$$
Geometric Distribution

Definition (Geometric Distribution)
In a series of Bernoulli trials (independent trials with constant probability $p$ of success), let the random variable $X$ denote the number of trials until the first success. Then, $X$ is a geometric random variable with parameter $p$ such that $0 < p < 1$ and the probability mass function of $X$ is

$$f(x) = (1 - p)^{x-1}p$$ for $x = 1, 2, 3, \ldots$

Example (Geometric random variable)
Let $X$ be a geometric random variable with $p = 0.25$. What is the probability that $X = 4$ (i.e. that the first success occurs on the 4th trial)?

**ANS:** For $X$ to be equal to 4, we must have had 3 failures, and then a success.

$$P(X = 4) =$$
Geometric Distribution

- The geometric distribution places mass at the counting numbers starting at 1 \( \{1, 2, 3, \ldots \} \). Though \( X \) can be any positive integer, the mass at the right-tailed numbers eventually gets quite small.

Comparison/Contrast with binomial distribution:

**Similar** They both have independent Bernoulli trials.

**Different** For a r.v. with a binomial distribution, we *know* how many trials we will have, there will be \( n \) trials, \( X \sim \text{Bin}(n, p) \) and \( X \in \{0, 1, 2, \ldots, n\} \). For a r.v. with a geometric distribution, we *do not know* how many trials we will have, \( X \sim \text{Geo}(p) \) and \( X \in \{1, 2, 3, \ldots \} \). We stop only when we get a success.
Geometric Distribution

- What does the distribution look like for $X \sim Geo(p)$?

- Is $f(x)$ a legitimate probability mass function?

\[ \sum_{x=1}^{\infty} (1 - p)^{x-1}p = p + (1 - p)p + (1 - p)^2p + \cdots = 1 \]

- Mathematically this works out because

\[ \sum_{x=0}^{\infty} a^x = \frac{1}{1 - a} \quad \text{for } 0 < a < 1 \]
Geometric Distribution

Definition (Mean and Variance for Geometric Distribution)

If \( X \) is a geometric random variable with parameter \( p \), then

\[
\mu = E(X) = \frac{1}{p} \quad \text{and} \quad \sigma^2 = V(X) = \frac{1-p}{p^2}
\]

Example (Weld strength)

A test of weld strength involves loading welded joints until a fracture occurs. For a certain type of weld, 80% of the fractures occur in the weld itself, while the other 20% occur in the beam. A number of welds are tested and the tests are independent.

Let \( X \) be the number of test at which the first beam fracture is observed.

1. Find \( P(X \geq 3) \)

{i.e. Find the probability that the first beam fracture happens on the third trial or later.}
1. Find $P(X \geq 3)$
{i.e. Find the probability that the first beam fracture happens on the third trial or later.}

ANS:

*Either a weld fracture or a beam fracture will occur on each Bernoulli trial. We’ll call a *success* a beam fracture. $X$ is a geometric random variable with $p = 0.20$. *

$$P(X \geq 3) = 1 - P(X < 3) \quad \text{\{complement\}}$$
$$= 1 - [P(X = 1) + P(X = 2)]$$
$$= 1 - [0.20 + (0.80)(0.20)]$$
$$= 0.64$$
Example (Weld strength, cont.)

2. Find $E(X)$ and $V(X)$ and the standard deviation of $X$.

**ANS:**

$E(X) = \frac{1}{p} = \frac{1}{0.20} = 5$

$\sigma^2 = V(X) = \frac{1-p}{p^2} = \frac{0.8}{(0.2)^2} = 20$

$\sigma = \sqrt{V(X)} = \sqrt{20} \approx 4.47$
Negative Binomial Distribution

- This distribution is similar to the geometric distribution, but now we’re interested in continuing the independent Bernoulli trials until \( r \) successes have been found (you must specify \( r \)).

Example (Weld strength, cont.)

Find the probability that the 3rd beam fracture (success) occurs on the 6th trial.

ANS: Recall, \( P(\text{success}) = P(\text{beam fracture}) = 0.2 \)

We want the probability that there were 2 successes somewhere within the first 5 trials, and the 6th trial was a success.

\[
\text{probability} = \left[ \binom{5}{2}(0.2)^2(0.8)^3 \right](0.2)
\]

\[
= \binom{5}{2}(0.8)^3(0.2)^3 = 10(0.8)^3(0.2)^3 = 0.04096
\]
Negative Binomial Distribution

**Definition (Negative Binomial Distribution)**

In a series of Bernoulli trials (independent trials with constant probability $p$ of success), let the random variable $X$ denote the number of trials until $r$ successes occur. Then, $X$ is a negative binomial random variable with parameters $0 < p < 1$ and $r = 1, 2, 3, \ldots$ and the probability mass function of $X$ is

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

for $x = r, r + 1, r + 2, \ldots$

Note that at least $r$ trials are needed to get $r$ successes.

- **NOTE:** the geometric distribution is a special case of the negative binomial distribution such that $r = 1$. 
Negative Binomial Distribution

A visual of the negative binomial distribution (given $p$ and $r$):

![Graph showing the negative binomial distribution with different values of $r$ and $p$.]
Definition (Mean and Variance for Negative Binomial Distribution)

If $X$ is a negative binomial random variable with parameters $p$ and $r$, then

$$\mu = E(X) = \frac{r}{p}$$

$$\sigma^2 = V(X) = \frac{r(1-p)}{p^2}$$

Example (Weld strength, cont.)

Let $X$ represent the number of trials until 3 beam fractures occur.

Then, $X$ follows a negative binomial distribution with parameters $p = 0.2$ and $r = 3$.

When will the 3rd beam fracture likely occur? On the 3rd trial? On the 4th trial?
1. Find $P(X = 4)$.

**ANS:**

2. Find the expected value of $X$ and the variance of $X$.

**ANS:**

$E(X) = \frac{3}{0.2} = 15$

$\sigma^2 = V(X) = \frac{3(0.8)}{(0.2)^2} = 60$ and $\sigma = \sqrt{60} \approx 7.746$
A pediatrician wishes to recruit 5 couples, each of whom is expecting their first child, to participate in a new natural childbirth regimen. Let \( p = P(\text{a randomly selected couple agrees to participate}) \).

If \( p = 0.15 \), what is the probability that 15 couples must be asked before 5 are found who agree to participate? That is, with \( S = \text{agrees to participate} \), what is the probability that 10 \( F \)'s occur before the fifth \( S \)?

**ANS:**
Let \( X \) represent the trial on which the 5\(^{th} \) success occurs.

\( X \) is a negative binomial r.v. with \( r = 5 \) and \( p = 0.15 \).

\[
P(X = 15) = f(15) = \binom{14}{4} (0.85)^{10} (0.15)^5 = 0.0150
\]
Comparison/Contrast with geometric distribution:

Similar The geometric is just a special case of the negative binomial when there is only one success, $r = 1$. 
The hypergeometric distribution has many applications in finite population sampling.

- Drawing from a relatively small population without replacement.

- When removing one object from the population of interest affects the next probability (this is in contrast to sequential trials for the binomial which had independent draws, where the probability of a success remained constant at $p$).

- When sequential draws are not independent, the probability depends on what occurred in earlier draws.
Hypergeometric Distribution

Example (Acceptance sampling)

Suppose a retailer buys goods in lots and each item in the lot can be either acceptable or defective.

If a lot has 25 items and 6 of them are defectives, what is the probability that a sample of size 10 has none that are defective?

We could apply our counting techniques to calculate this, but here, we’ll use the general formula for the hypergeometric probability distribution...
Hypergeometric Distribution

- Sampling without replacement in a small finite population

**Definition (Hypergeometric Distribution)**

A set of $N$ objects contains:

- $K$ objects classified as successes
- $N - K$ objects classified as failure

A sample of size $n$ objects is selected randomly (without replacement) from the $N$ objects, where $K \leq N$ and $n \leq N$. Let the random variable $X$ denote the number of successes in the sample. Then $X$ is a hypergeometric random variable and

$$f(x) = \frac{\binom{K}{x} \binom{N - K}{n - x}}{\binom{N}{n}} = \frac{\text{#successes}}{x} \cdot \frac{\text{#failures}}{n - x}$$

for $x = \max\{0, n - (N - K)\}^*$ to $\min\{K, n\}^*$
Hypergeometric Distribution

Example (Acceptance sampling, cont.)

Set-up the hypergeometric distribution...

Information we’re given:

There are 25 objects altogether, and 6 are defectives.
We also know that we are going to choose or ‘sample’ 10 objects.

We’re going to calculate the probability of choosing 0 defects, so we’ll label ‘getting a defect’ as a success (it’s just a label).

A set of $N = 25$ objects contains:

\[ K = 6 \] objects classified as successes
\[ N - K = 19 \] objects classified as failure

And $n = 10$ (we’re choosing 10 objects).
Example (Acceptance sampling, cont.)

What is the probability that a sample of size 10 has no defectives?

**ANS:** Let \( X \) be the number of defectives (or successes) found in a sample of size \( n=10 \). \( X \) is a hypergeometric random variable because we’re drawing without replacements from a small population.

We wish to calculate \( P(X = 0) \).

The probability mass function (PMF) for \( X \) is shown below:

\[
P(X = x) = f(x) = \binom{6}{x} \binom{19}{10-x} \binom{25}{10} 
\text{ for } x \in \{0, 1, 2, 3, 4, 5, 6\}
\]
Hypergeometric Distribution

Example (Acceptance sampling, cont.)

\[
P(X = 0) = f(0) = \frac{\binom{6}{0} \binom{19}{10}}{\binom{25}{10}} = 0.0283
\]

You can use the PMF to generate the probability for any relevant \( x \)... 

\[
P(X = x) = f(x) = \frac{\binom{6}{x} \binom{19}{10 - x}}{\binom{25}{10}}
\]

Notice that even though you're drawing \( n=10 \) it is impossible to draw any more than 6 successes (there's only 6 altogether).
POSSIBLE VALUES FOR $X$ if $X$ is a hypergeometric random variable:

highest possible** $= \min\{K, n\} = \min\{\#successes, n\}$

The largest number of successes you can have is either $K$ (the # of successful objects present) or the number of trials (whichever of these is a smaller number, thus, $\min\{K, n\}$).

lowest possible* $= \max\{0, n - (N - K)\} = \max\{0, \#chosen - \#failures\}$

If there are only a few failure objects available, then you’ll have to choose some successes simply by default. If there’s only a few failures, the smallest possible number of successes is not 0 and specifically it is $n - (\#\text{ of failures in the total pool}) = n - (N - K)$. If the number of failures is $\geq n$, then you can have all failures, and the number of successes $X$ can be 0.

So, the smallest number of successes you can have depends on the number of non-successes $(N - K)$ and how many you’re choosing $n$. 
Example (Acceptance sampling, cont.)

Continuing the example...

\[ N = 25 \]
\[ K = 6 \] \{successes\}
\[ N - K = 19 \] \{failures\}
\[ n = 10 \] \{number chosen\}

Minimum possible value of \( X \)? ___0___

\[ \text{max}\{0, n - (N - K)\} = \text{max}\{0, 10 - 19\} = \text{max}\{0, -9\} \]

Maximum possible value of \( X \)? ___6___

\[ \text{min}\{K, n\} = \text{min}\{6, 10\} \]

For the hypergeometric, you want to be able to recognize that you have a finite pool you’re drawing from without replacement (i.e. not independent sequential draws), and be able to compute the probabilities using the probability mass function.
Hypergeometric Distribution

Definition (Mean and Variance for Hypergeometric Distribution)

If $X$ is a negative hypergeometric random variable with parameters $N$, $K$ and $n$, then

$$\mu = E(X) = n \left( \frac{K}{N} \right)$$

$$\sigma^2 = V(X) = n \left( \frac{K}{N} \right) \cdot \left( 1 - \frac{K}{N} \right) \cdot \left( \frac{N-n}{N-1} \right)$$
Example (Acceptance sampling, cont.)

Find the $E(X)$ and $V(X)$.

**ANS:**

$n = 10, N = 25, K = 6$

$\mu = E(X) = 10 \cdot \frac{6}{25} = 2.4$

$\sigma^2 = V(X)$

$= 10 \cdot \frac{6}{25} \cdot \frac{19}{25} \cdot \frac{15}{24}$

$= 1.14$
Hypergeometric Distribution

Comparison/Contrast with binomial, geometric, and negative binomial:

Similar They all consider successes and failures for a set of trials.

Different In the hypergeometric, we do not have a constant probability of a success $p$ in sequential trials. The probability of a success on one trial depends on what was chosen in previous trials. This is because the population from which you are drawing is finite and relatively small.

The other three random variables listed above have independent Bernoulli trials whereas the hypergeometric does not.
Section 3.8 Poisson Distribution

This discrete distribution will not be on Midterm Exam 1, but it will be on Midterm Exam 2 instead.