STA 247
Probability with Computer Applications

Professor K. H. Wong

Week 4 - Topic C

Hypergeometric Distribution

Example 7 A lot of 5000 items contains 10% defective items. If a random sample of 5 items is tested, find the probability of observing at least one defective item.

Hypergeometric Distribution

Example 7 A lot of 5000 items contains 10% defective items. If a random sample of 5 items is tested, find the probability of observing at least one defective item.

- We have a sample size $n = 5$ and a probability of "success" $p = 0.1$ so it's tempting to use the binomial distribution here to compute the probability.

Hypergeometric Distribution

Example 7 A lot of 5000 items contains 10% defective items. If a random sample of 5 items is tested, find the probability of observing at least one defective item.

- We have a sample size $n = 5$ and a probability of "success" $p = 0.1$ so it's tempting to use the binomial distribution here to compute the probability.
- However, this scenario is not one of **independent samples** which is the foundation for binomial distribution.

Hypergeometric Distribution

Example 7 A lot of 5000 items contains 10% defective items. If a random sample of 5 items is tested, find the probability of observing at least one defective item.

- We have a sample size $n = 5$ and a probability of "success" $p = 0.1$ so it's tempting to use the binomial distribution here to compute the probability.
- However, this scenario is not one of **independent samples** which is the foundation for binomial distribution.
- In essence, this lot has 500 defective items and 4500 functional items. When you select a random sample of 5, you are selecting **without replacement** which will change the probability of selecting a defective item with each trial.

Example 7 A lot of 5000 items contains 10% defective items. If a random sample of 5 items is tested, find the probability of observing at least one defective item.

- We have a sample size $n = 5$ and a probability of "success" $p = 0.1$ so it's tempting to use the binomial distribution here to compute the probability.
- However, this scenario is not one of **independent samples** which is the foundation for binomial distribution.
- In essence, this lot has 500 defective items and 4500 functional items. When you select a random sample of 5, you are selecting **without replacement** which will change the probability of selecting a defective item with each trial.
- On the other hand, a sample of 5 is negligible compared to the size of the lot and the number of defectives, so even without replacement, it will not change the probability of selecting a defective item much.

Hypergeometric Distribution

Example 7 A lot of 5000 items contains 10% defective items. If a random sample of 5 items is tested, find the probability of observing at least one defective item.

- We have a sample size $n = 5$ and a probability of "success" $p = 0.1$ so it's tempting to use the binomial distribution here to compute the probability.
- However, this scenario is not one of **independent samples** which is the foundation for binomial distribution.
- In essence, this lot has 500 defective items and 4500 functional items. When you select a random sample of 5, you are selecting **without replacement** which will change the probability of selecting a defective item with each trial.
- On the other hand, a sample of 5 is negligible compared to the size of the lot and the number of defectives, so even without replacement, it will not change the probability of selecting a defective item much.
- You can reasonably approximate the probability by using the binomial distribution.

Example 7 A lot of 5000 items contains 10% defective items. If a random sample of 5 items is tested, find the probability of observing at least one defective item.

Hypergeometric Distribution

Example 7 A lot of 5000 items contains 10% defective items. If a random sample of 5 items is tested, find the probability of observing at least one defective item.

- Let $D =$ Number of defectives in the random sample
- Probability of success $p = 0.1$, probability of failure $q = 1 - 0.1 = 0.9$

  $P(D \geq 1) = 1 - P(D < 1) = 1 - P(D = 0)$

  $P(D \geq 1) = 1 - \binom{5}{0}(0.1)^0 \cdot (0.9)^5 = 1 - 0.59049 = 0.40951$

  $\therefore$ 41% chance of selecting at least one defective item.

Hypergeometric Distribution

Example 8 What if instead we have a lot of 50 items, 10% of which are known to be defective. Find the probability of observing at least one defective item.

Hypergeometric Distribution

Example 8 What if instead we have a lot of 50 items, 10% of which are known to be defective. Find the probability of observing at least one defective item.

- Identical problem to **example 7** with one caveat: Your sample size vs. lot and defective sizes are no longer negligible.

Hypergeometric Distribution

Example 8 What if instead we have a lot of 50 items, 10% of which are known to be defective. Find the probability of observing at least one defective item.

- Identical problem to **example 7** with one caveat: Your sample size vs. lot and defective sizes are no longer negligible.
- You now have 5 defective items, 45 functional items, and a sample size $n = 5$. Removing 5 items from a lot of 50 will significantly alter the probabilities of selecting a defective item in each stage. Each trial is *dependent* on the previous.

Example 8 What if instead we have a lot of 50 items, 10% of which are known to be defective. Find the probability of observing at least one defective item.

- Identical problem to **example 7** with one caveat: Your sample size vs. lot and defective sizes are no longer negligible.
- You now have 5 defective items, 45 functional items, and a sample size $n = 5$. Removing 5 items from a lot of 50 will significantly alter the probabilities of selecting a defective item in each stage. Each trial is *dependent* on the previous.
- Instead, we can approach this as a *counting* problem. How many ways can we select 1, 2, 3, 4, 5 defective items out of all the ways we can select 5 items from the lot of 50?

Example 8 What if instead we have a lot of 50 items, 10% of which are known to be defective. Find the probability of observing at least one defective item.

- Identical problem to **example 7** with one caveat: Your sample size vs. lot and defective sizes are no longer negligible.
- You now have 5 defective items, 45 functional items, and a sample size $n = 5$. Removing 5 items from a lot of 50 will significantly alter the probabilities of selecting a defective item in each stage. Each trial is *dependent* on the previous.
- Instead, we can approach this as a *counting* problem. How many ways can we select 1, 2, 3, 4, 5 defective items out of all the ways we can select 5 items from the lot of 50?
- Or using the indirect approach, how many ways can we select 0 defective items out of all the ways we can select 5 items from the lot of 50?

Hypergeometric Distribution

Example 8 What if instead we have a lot of 50 items, 10% of which are known to be defective. Find the probability of observing at least one defective item.

Hypergeometric Distribution

Example 8 What if instead we have a lot of 50 items, 10% of which are known to be defective. Find the probability of observing at least one defective item.

$$P(D \geq 1) = 1 - P(D < 1) = 1 - P(D = 0)$$

$$P(D \geq 1) = 1 - \frac{\binom{5}{0}\binom{45}{5}}{\binom{50}{5}} = 1 - 0.576639 = 0.42336$$

- $\therefore$ 42% chance of selecting at least one defective item.

Hypergeometric Distribution

key points is that because there is no replacement, the trails are dependent

Hypergeometric Distribution

Suppose you have a pool of N objects that can be partitioned into 2 (or more groups) by some characteristic. Suppose there are k objects of type A and $N-k$ objects of type B. In a random sample of size n (**without replacement**) from this pool of N objects, let X denote the random variable for the number of objects of type A that is selected.

$$P(X=x)=\frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$$

similar to E(x) = np, for binomial distribution, here k/N is probability of success on one draw

X has an expected value $E[X]=n\cdot\frac{k}{N}$ and variance

$V(X)=n\cdot\frac{k}{N}(1-\frac{k}{N})(\frac{N-n}{N-1})$ correction factor for dependent samples, the rest is similar to V(X) = np(1-p) for binomial

When N is large, the probability of success change little with each random selection; binomial distribution approximate hypergeometric probabilities well. However since most populations are finite, many of binomial application in which sampling is without replacement involve finite populations so the distribution is actually hypergeometric