Section 3.7: Hypergeometric and Negative Binomial Distributions

1 The Hypergeometric Distribution

The geometric probability distribution looks for the first success where selections are made with replacement (or the sample size is less than 5% of the population size). The hypergeometric distribution addresses the experiments where selections are made without replacement.

Definition 1 *An experiment consisting of repeated trials has a* ***Hypergeometric Probability Distribution*** *if and only if*

1. *the population is finite of size N;*
2. *there are only two outcomes, success or failure, and the population contains M successes;*
3. *each sample of size n is equally likely.*

Knowing an experiment has a hypergeometric distribution provides the following formulae.

Theorem 2 *If an experiment has a hypergeometric probability distribution with parameters n, N and M then*

1. *the probability of exactly x successes is $p(x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$;*
2. *the mean or expected value of the number of successes*

$$E(X) = n * \frac{M}{N}$$

which by letting $p = \frac{M}{N}$ transforms into

$$E(X) = np$$

3. *and the variance of the number of successes is*

$$V(X) = \left(\frac{N-n}{N-1}\right) n \frac{M}{N} \left(1 - \frac{M}{N}\right)$$

which by letting $p = \frac{M}{N}$ transforms into

$$V(X) = \left(\frac{N-n}{N-1}\right) np(1-p).$$

Proof. Note that since selections are made without replacement, we are choosing unordered subsets.

1. There are clearly $\binom{N}{n}$ different ways to select a sample of size n from a population of size N. With M successes, we can select x of those in $\binom{M}{x}$ ways. If there are M successes in the population then there must be $N - M$ failures in the population. We need to pick $n - x$ of those for the sample. This can be done in $\binom{N-M}{n-x}$ ways. Thus, the probability of exactly x successes is $p(x) = \dfrac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$.

2. According to the textbook, the moment generating function for the hypergeometric probability distribution is "more trouble than it worth here."

3. Ditto from 2.

■

Example 3 *Fred's horror movie collection consists of 10 films that have a sense of humor and 8 that do not. Fred plans a Saturday night triple feature and randomly selects three horror films from his collection. What is the probability that Fred picks two films that have a sense of humor and 1 that does not? Here* $N = 18$, $n = 3, M = 10$ *and* $x = 2$. *So,*

$$\begin{aligned} p(x) &= \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} \\ &\quad \textit{So, } p(2) \\ &= \frac{\binom{10}{2}\binom{8}{1}}{\binom{18}{3}} \\ &= 0.441\,18. \end{aligned}$$

Remark 4 *I think it is much easier to understand the hypergeometric formula from the perspective of success and failure than using so many variables. Your mileage may vary. Furthermore we could consider picking a film with no sense of humor as a success and compute* $p(1)$. *This yields* $p(1) = \frac{\binom{8}{1}\binom{10}{2}}{\binom{18}{3}} = 0.441\,18$. *Naturally this is the same value since multiplication is a commutative operation.*

Example 5 *When Fred randomly selects three horror films from his collection what is* $E(X), V(X)$ *and* σ *for the number of films that have a sense of humor? Note that* $p = \frac{M}{N} = \frac{10}{18}$ *for selecting a film with a sense of humor. So,*

$$\begin{aligned} E(X) &= np \\ &= 3 * \frac{10}{18} \\ &= 1.666\,7 \end{aligned}$$

and

$$\begin{aligned} V(X) &= \left(\frac{N-n}{N-1}\right) np(1-p) \\ &= \left(\frac{18-3}{18-1}\right) * 3 * \frac{10}{18} * (1 - \frac{10}{18}) \\ &= 0.653\,59 \end{aligned}$$

Thus, $\sigma = \sqrt{0.653\,59} = 0.808\,45$.

Example 6 *Back to Jason's CD collection that consists of five different rock CDs, three different jazz CDs, two different blues CDs, two different classical CDs and a single folk CD. Jason is planning a trip and randomly selects four CDs. What is the probability that Jason selects three rock CDs? We now see this as a hypergeometric problem where selecting a rock CD is considered a success. So,*

$$\begin{aligned} p(x) &= \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} \\ & \textit{So, } p(3) \\ &= \frac{\binom{5}{3}\binom{8}{1}}{\binom{13}{4}} \\ &= 0.111\,89. \end{aligned}$$

What are $E(X), V(X)$ *and* σ *for the number of rock CDs that Jason selects?* $E(X) = np = 4 * \frac{5}{13} = 1.538\,5$.

$$\begin{aligned} V(X) &= \left(\frac{N-n}{N-1}\right) np(1-p) \\ &= \left(\frac{13-4}{13-1}\right) 4 * \frac{5}{13}(1 - \frac{5}{13}) \\ &= 0.710\,06 \end{aligned}$$

Hence, $\sigma = \sqrt{0.710\,06} = 0.842\,65$.

Example 7 *Jason also packs his 6 can cooler to capacity for the road trip at random from a fridge with 27 beers and 13 sodas.*

1. What is the probability that he randomly picks three beers and three sodas? $p(3) = \frac{\binom{27}{3}\binom{13}{3}}{\binom{40}{6}} = 0.217\,94$.

2. What is the probability that the number of beers is larger than the number of sodas?

$$\begin{aligned} & p(4) + p(5) + p(6) \\ &= \frac{\binom{27}{4}\binom{13}{2}}{\binom{40}{6}} + \frac{\binom{27}{5}\binom{13}{1}}{\binom{40}{6}} + \frac{\binom{27}{6}\binom{13}{0}}{\binom{40}{6}} \\ &= 0.707\,17. \end{aligned}$$

Remark 8 *Note that the binomial probability distribution and the hypergeometric distribution have the same expected value np. The $V(X)$ of the two distributions differ only by $\left(\frac{N-n}{N-1}\right)$. This is referred to as the finite population correction factor. Note that when n is small compared to N, $\left(\frac{N-n}{N-1}\right)$ is approximately 1. Mathematically we see that when n is small compared to N it is very reasonable to use the binomial distribution to approximate the hypergeometric distribution (assume selections are made with replacement.*

2 The Negative Binomial Distribution

Definition 9 *An experiment consisting of repeated trials has a* ***Negative Binomial Probability Distribution*** *if and only if*

1. *there are only two outcomes for each trial (success or failure) and the probability p of success is fixed;*
2. *the trials are independent;*
3. *the experiment continues until there are r successes.*

The random variable X is the number of failures that precede the r^{th} success. Knowing an experiment has a negative binomial distribution provides the following formulae.

1. The probability of x failures before the r^{th} success is $p(x) = \binom{x+r-1}{r-1}p^r(1-p)^x$;Note that $\binom{x+r-1}{x} = \binom{x+r-1}{r-1}$;
2. $M_X(t) = \frac{p^r}{[1-e^t(1-p)]^r}$;
3. $E(X) = \frac{r(1-p)}{p}$;
4. $V(X) = \frac{r(1-p)}{p^2}$.

1. If we have x failures before the r^{th} success then we have $x + r$ trials. Furthermore, the $x + r^{th}$ is a success. The remaining $r - 1$ successes can occur anywhere is the first $x + r - 1$ trials. Thus, we need to pick which of those trials yields the $r - 1$ successes. We can do so in $\binom{x+r-1}{r-1}$ ways. Then we must attain those r successes and fail the remaining x times. Hence, $p(x) = \binom{x+r-1}{r-1}p^r(1-p)^x$.
2. Once again we use the binomial theorem that states $(a+b)^n = \sum_{x=0}^{n} \binom{n}{x} a^x b^{n-x}$

for all integers n and values of a and b. Recall that

$$\begin{aligned}
\binom{n}{x} &= \frac{n!}{x!(n-x)!} \\
&= \frac{n(n-1)(n-2)...(n-x+1)}{x!} \\
\text{letting } n &= -r \text{ we get} \\
\binom{-r}{x} &= \frac{(-r)!}{x!(-r-x)!} \\
&= \frac{-r(-r-1)(-r-2)...(-r-x+1)}{x!} \\
&= (-1)^x \frac{(r+x-1)...(r+2)(r+1)r}{x!} \\
&= (-1)^x \frac{(r+x-1)...(r+2)(r+1)r}{x!} * \frac{(r-1)!}{(r-1)!} \\
&= (-1)^x \binom{x+r-1}{r-1} \\
& \quad \text{since } (x+r-1)-x \\
&= (r-1)
\end{aligned}$$

By the binomial theorem and letting $a = 1, b = -b$ and $n = -r$ we have

$$\begin{aligned}
& (1-b)^{-r} \\
= & \sum_{k=0}^{-r} \binom{-r}{x} (-b)^x 1^{-r-x} \\
= & \sum_{k=0}^{-r} (-1)^x \binom{x+r-1}{r-1} (-b)^x \\
= & \sum_{k=0}^{-r} (-1 * -b)^x \binom{x+r-1}{r-1} \\
= & \sum_{k=0}^{-r} b^x \binom{x+r-1}{r-1}.
\end{aligned}$$

Thus,

$$\begin{aligned}
M_x(t) &= \sum_{x\in X} e^{tx}p(x) \\
&= \sum_{x\in X} e^{tx}\binom{x+r-1}{r-1}p^r(1-p)^x \\
&= p^r\sum_{x\in X} e^{tx}\binom{x+r-1}{r-1}(1-p)^x \\
&= p^r\sum_{x\in X} \binom{x+r-1}{r-1}\left[e^t(1-p)\right]^x \\
&= p^r * (1-(\left[e^t(1-p)\right]))^{-r} \\
&= \frac{p^r}{(1-([e^t(1-p)]))^r}.
\end{aligned}$$

3. $E(X) = M_x^1(0)$. Remember t is our variable. With

$$\begin{aligned}
M_x(t) &= p^r(1-(\left[e^t(1-p)\right]))^{-r} \text{ we get that} \\
M_x^1(t) &= p^r(-r)(1-(\left[e^t(1-p)\right]))^{-r-1} * (-e^t(1-p))
\end{aligned}$$

and

$$\begin{aligned}
M_x^1(0) &= p^r(-r)(1-(1-p))^{-r-1} * (-1(1-p)) \\
&= p^r(r)(p)^{-r-1} * (1-p) \\
&= \frac{p^r(r)(1-p)}{p^{r+1}} \\
&= \frac{r(1-p)}{p}
\end{aligned}$$

4. We let $V(X) = \frac{r(1-p)}{p^2}$ stand without proof.

Example 10 *A pair of fair dice is rolled.*

1. What is the likelihood we roll the dice 13 times before seeing a sum of 7 twice? Using $p(x) = \binom{x+r-1}{r-1}p^r(1-p)^x$, $p(11) = \binom{11+2-1}{2-1} * \frac{1}{6}^2 * (1-\frac{1}{6})^{11} = 4.4863 \times 10^{-2} = 0.0448$

2. What is the expected number of failures before rolling two sums of 7? $E(X) = \frac{r(1-p)}{p} = \frac{2(1-\frac{1}{6})}{\frac{1}{6}} = 10.0$

3. What is the variance and standard deviation of the number of failures before rolling two sums of 7? $V(X) = \frac{2(1-\frac{1}{6})}{\frac{1}{6}^2} = 60.0$ and $\sigma = \sqrt{60} = 7.7460$

4. What is the likelihood we roll the dice no more than 5 times before seeing a sum of 7 twice?

$$\begin{aligned}
& \sum_{x=0}^{3} p(x) \\
= & \sum_{x=0}^{3} \binom{x+r-1}{r-1} p^r (1-p)^x \\
= & \sum_{x=0}^{3} \binom{x+2-1}{2-1} \frac{1}{6}^2 (1-\frac{1}{6})^x \\
= & \sum_{x=0}^{3} \binom{x+1}{1} \frac{1}{36} (\frac{5}{6})^x \\
= & \sum_{x=0}^{3} (x+1) \frac{1}{36} (\frac{5}{6})^x \\
= & (1\frac{1}{36}(\frac{5}{6})^0 + (2)\frac{1}{36}(\frac{5}{6})^1 + (3)\frac{1}{36}(\frac{5}{6})^2 + (4)\frac{1}{36}(\frac{5}{6})^3 \\
= & 0.196\,24
\end{aligned}$$

5. What is the likelihood we roll the dice at least 6 times before seeing a sum of 7 twice? $1 - \sum_{x=0}^{3} p(x) = 1 - 0.196\,24 = 0.803\,76$.