nurdan.kar@bilkent.edu.tr

3.7. Other discrete distributions

In this section, we introduce some discrete distributions that are related to the binomial distribution.

The hypergeometric distribution

We have the following assumptions:

1. The population has a total of N objects (it is a finite population)
2. Each object is either a success or a failure, and there are M successes in total.
3. A sample of n objects is chosen without replacement, and every possible subset of size n is equally likely.

In this distribution, the rv

X: the number of successes in the sample space

and the pmf is denoted by $h(x; n, M, N)$.

Proposition. Let N, M, and n be positive integers with $\max(0, n-N+M) \leq x \leq \min(n, M)$.

$$P(X=x) = h(x; n, M, N) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

is said to be a hypergeometric pmf.

Example. In 500 independent calculations a scientist has made 25 errors. If a second scientist checks seven of these calculations randomly, what is the probability that he detects two errors? Assume that the second scientist will definitely find the error of a false calculation.

Solution. X: the numb. of errors found by the second scientist.

Then X is hypergeometric with $N=500$, $M=25$, and $n=7$. We wish to find $P(X=2)$.

$$P(X=2) = h(2; 7, 25, 500) = \frac{\binom{25}{2}\binom{500-25}{7-2}}{\binom{500}{7}} \approx .04.$$

Proposition. The mean and variance of the hypergeometric rv X having pmf $h(x; n, M, N)$ are

$$E(X) = n \cdot \frac{M}{N}, \quad V(X) = \left(\frac{N-n}{N-1}\right) \cdot n \frac{M}{N}\left(1 - \frac{M}{N}\right).$$

Here, $\frac{M}{N}$ is the proportion of successes in the population. Replacing it by p in $E(X)$ and $V(X)$ gives

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

Here, $\left(\frac{N-n}{N-1}\right) < 1$ is called the finite population correction factor.

Example. In an assembly-line production of industrial robots, gearbox assemblies can be installed in 1 minute each if holes have been properly drilled in the boxes and in 10 minutes each if the holes must be redrilled. 20 gearboxes are in stock, and it is assumed that two will have improperly drilled holes. 5 gearboxes are to be randomly selected from the 20 available for installation in the next five robots in line.

a) Find the probability that all five boxes will fit properly.

b) Find $E(T)$, $V(T)$, and $SD(T)$ of the time it takes to install these five boxes.

Solution. $N = 20$, $M = 2$, $n = 5$

Suppose X denote the numb. of boxes with improperly drilled holes in the sample of five.

a) $P(\text{all five boxes will fit properly}) = P(X=0) = \dfrac{\binom{2}{0}\binom{18}{5}}{\binom{20}{5}} = \dfrac{(1)(8,568)}{15,504} = .55$

b) The total time for installation is $T = \overbrace{10X}^{\text{improper boxes take 10 min to install}} + \underbrace{(5-X)}_{\text{proper boxes take 1 min.}} = 9X + 5$

We wish to find $E(T)$, that is $E(T) = 9 . E(X) + 5$. Let us find $E(X)$ and $V(X)$ first.

$E(X) = n . \frac{M}{N} = 5 . \frac{2}{20} = .5$, $V(X) = \left(\frac{N-n}{N-1}\right) n . \frac{M}{N}\left(1 - \frac{M}{N}\right) = \left(\frac{20-5}{20-1}\right) . 5 . \frac{2}{20} . \left(1 - \frac{2}{20}\right) = .355$

Then,

$E(T) = 9(.5) + 5 = 9.5$, $V(T) = V(9X+5) = 9^2 V(X) = 9^2(.355) = 81(.355) = 28.755$

$V(aX+b) = a^2 V(X)$

$SD(X) = \sqrt{28.755} = 5.4\text{min}$

The negative binomial and geometric distributions

The binomial rv X is the number of successes with the fixed numb. of trials. However, the negative binomial fixes the number of successes and letting the number of trials be random.

The negative binomial distribution has the following conditions:

1. The experiment has a sequence of independent trials.
2. Each trial can result in either a success or a failure.
3. $P(\text{success}) = p$ is constant from trial to trial.
4. The experiment continues until a total of r successes have been observed, where r is a specified positive integer.

In this distribution, the rv

X : the number of trials required to achieve the r-th success,

and the pmf is denoted by

$$nb(x; r, p) = P(X = x) = \overbrace{P(r-1 \text{ successes on the first } x-1 \text{ trials})}^{\text{binomial prob}} \cdot \overbrace{P(\text{success})}^{p}$$

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

Then, we have the following proposition.

Proposition. The pmf of the negative binomial rv X with parameters $r =$ desired number of successes and $p = P(S)$ is

$$nb(x; r, p) = \binom{x-1}{r-1} p^{r}(1-p)^{x-r}, \quad x = r, r+1, r+2, \dots .$$

Note that a negative binomial pmf with parameters $(1, p)$ is geometric. Namely,

$$nb(x; 1, p) = p(1-p)^{x-1}, \quad x = 1, 2, \dots .$$

is called the geometric distribution. In this distribution, rv X is the number of trials required to achieve one success. We have

1. $nb(x; 1, p) \geq 0$, 2. $\sum nb(x; 1, p) = p + p(1-p) + p(1-p)^2 + \dots = \frac{p}{1-(1-p)} = 1$.

Proposition. If X is a negative binomial r.v with parameters r and p, then

$$E(X) = \frac{r}{p}, \quad V(X) = \frac{r(1-p)}{p^2}, \quad M_X(t) = \left(\frac{pe^t}{1-(1-p)e^t}\right)^r$$

Example. Sharon and Jane play a series of backgammon games until one of them wins five games. Suppose that the games are independent and the probability that Sharon wins a game is .58.

a) Find the probability that the series ends in seven games.

b) If the series ends in seven games, what is the probability that Sharon wins?

Solution. a) X: the numb. of games until Sharon wins five games.

Y: " " " " Jane " "

Sharon needs to achieve 5 wins (successes). Then, this is negative binomial. Thus, X and Y are negative binomial with parameters $(5, .58)$ and $(5, .42)$, respectively. The prob. that the series ends in seven games:

$$P(X=7) + P(Y=7) = \binom{6}{4}(.58)^5(.42)^2 + \binom{6}{4}(.42)^5(.58)^2$$

$$\approx .17 + .066$$

$$\approx .24.$$

b) A: the event that Sharon wins

B: the event that the series ends in seven games

We wish to find $P(A|B)$. By the def. of conditional prob.,

$$P(A|B) = \frac{P(A\cap B)}{P(B)} = \frac{P(X=7)}{P(X=7)+P(Y=7)} \approx \frac{.17}{.24} \approx .71.$$

Alternative definition of the negative binomial distribution

In the literature, it is also possible to see the "number of failures preceding the rth success" called negative binomial. This is written by $X-r$ in our notation. Similarly, geo. dist. is sometimes defined using the number of failures preceding the first success in a sequence of independent and identical trials.