3.5 Hypergeometric and Negative Binomial Distributions Distributions

The Hypergeometric Distribution

Example 1. A lot consists of N =10 articles of which M = 6 are good (S) and N – M = 4 are defective (F). n=5 articles are selected from lot at random and without replacement.

1. Find the probability that exactly x = 3 of the selected are good.
2. Find the probability that exactly x of the selected are good, x = 0, 1, 2, 3, 4, 5

F	F	F	F		
S	S	S	S	S	S

Let X be the number of good articles in a sample of 5.

1. $P(X = 3) = \frac{\binom{6}{3}\binom{4}{2}}{\binom{10}{5}}$ 2. $P(X = x) = \frac{\binom{6}{x}\binom{4}{5-x}}{\binom{10}{5}}$, x = 1, 2, 3, 4, 5, and $P(X = 0) = 0$

The assumptions leading to the hypergeometric distribution are as follows:

1. The population or set to be sampled consists of N individuals, objects, or elements (a *finite* population).
2. Each individual can be characterized as a success (S) or a failure (F), and there are M successes in the population.
3. A sample of n individuals is selected without replacement in such a way that each subset of size n is equally likely to be chosen.

The random variable of interest is X = the number of S's in the sample. The probability distribution of X depends on the parameters n, M, and N, so we wish to obtain $P(X = x) = h(x; n, M, N)$.

PROPOSITION

If X is the number of S's in a completely random sample of size n drawn from a population consisting of M S's and $(N - M)$ F's, then the probability distribution of X, called the **hypergeometric distribution,** is given by

$$P(X = x) = h(x; n, M, N) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} \tag{3.15}$$

for x an integer satisfying $\max(0, n - N + M) \leq x \leq \min(n, M)$.

Restrictions for values of x: If we select *n* individuals from a population that consists of M successes and N-M failures and x is the number of successes in a sample, then $x \leq M$, and the number of failures in a sample $n - x \leq N - M$ and hence $x \geq n - N + M$.

PROPOSITION

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

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

The Negative Binomial Distribution

Example 2. We roll a die and success is if it turns "six"; here p = P(S) = 1/6.

1. What is the probability that the first time the success occurs at the $x=4^{th}$ trial?
2. What is the probability that we need to roll a die 10 times to get the three successes?

1. Let Y_1 be the number of the 1^{st} successful trial. Then $P(Y_1 = 4) = P(FFFS) = \left(\frac{5}{6}\right)^3\left(\frac{1}{6}\right)$

   In general $P(Y_1 = y) = P(FF \ldots FS) = \left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^{y-1}$

2. Let now Y_3 be the number of the 3rd successful trial. Then

$$P(Y_3 = 10) = P(2\ succeses\ in\ 9\ trials\ AND\ a\ success\ in\ 10th\ trial\ ) = \binom{9}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^7 \times \left(\frac{1}{6}\right)$$

$$P(Y_3 = 10) = \binom{9}{2}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^7$$

In general the probability that the r-th success occurs at the trial number y, is

$$P(Y_r = y) = \binom{y-1}{r-1}(p)^r(1-p)^{y-r}$$

So if X = the number of failures that precede r-th success, then

$$P(X = x) = P(Y_r = x + r) = \binom{x+r-1}{r-1}(p)^r(1-p)^x = nb(x; r, p)$$

Assumptions:

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

The random variable of interest is X = the number of failures that precede the rth success; X is called a **negative binomial random variable** because, in contrast to the binomial rv, the number of successes is fixed and the number of trials is random.

PROPOSITION

The pmf of the negative binomial rv X with parameters r = number of S's and $p = P(S)$ is

$$nb(x; r, p) = \binom{x + r - 1}{r - 1} p^r (1 - p)^x \qquad x = 0, 1, 2, \ldots$$

PROPOSITION

If X is a negative binomial rv with pmf $nb(x; r, p)$, then

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

Special case r = 1 is called the **Geometric Distribution.**

Let X be number of failures that precede the 1st success, and Y_1 be the number of the first successful trial. Then

$$P(X = x) = P(Y_1 = x + 1) = (1 - p)^x p, \quad x = 0, 1, 2, \ldots$$

$$E(X) = \frac{1-p}{p}, \; V(X) = \frac{1-p}{p^2}$$

EXERCISES 3.5

68. An electronics store has received a shipment of 20 table radios that have connections for an iPod or iPhone. Twelve of these have two slots (so they can accommodate both devices), and the other eight have a single slot. Suppose that six of the 20 radios are randomly selected to be stored under a shelf where the radios are displayed, and the remaining ones are placed in a storeroom. Let X = the number among the radios stored under the display shelf that have two slots.

a. What kind of a distribution does X have (name and values of all parameters)?

b. Compute $P(X = 2)$, $P(X \le 2)$, and $P(X \ge 2)$.

c. Calculate the mean value and standard deviation of X.

Answers:

a. hypergeometric (n=6, M=12, N=20)

b. 0.119, 0.137, 0.982

c. 3.6, 1.061

74. A second-stage smog alert has been called in a certain area of Los Angeles County in which there are 50 industrial firms. An inspector will visit 10 randomly selected firms to check for violations of regulations.

a. If 15 of the firms are actually violating at least one regulation, what is the pmf of the number of firms visited by the inspector that are in violation of at least one regulation?

b. If there are 500 firms in the area, of which 150 are in violation, approximate the pmf of part (a) by a simpler pmf.

c. For X = the number among the 10 visited that are in violation, compute $E(X)$ and $V(X)$ both for the exact pmf and the approximating pmf in part (b).

Answers:

a. h(x; 10,15,50)

b. h(x; 10,150,500) ≈b(10,0.3)

c. exact: .3, 2.06;

approximate: .3, 2.1

75. Suppose that $p = P(\text{male birth}) = .5$. A couple wishes to have exactly two female children in their family. They will have children until this condition is fulfilled.

a. What is the probability that the family has x male children?

b. What is the probability that the family has four children?

c. What is the probability that the family has at most four children?

d. How many male children would you expect this family to have? How many children would you expect this family to have?

Answers:

a. $(x+1)p^x(1-p)^2 = (x+1)0.5^{x+2}$

b. 0.1875

c. 0.6875

d. 2,4