MODULE 4

SOME SPECIAL DISCRETE DISTRIBUTIONS AND THEIR PROPERTIES

LECTURE 18

Topics

4.2 NEGATIVE BINOMIAL DISTRIBUTION

4.3 THE HYPERGEOMETRIC DISTRIBUTION

Definition 1.2

Let T be a random variable of discrete type with support $S_T = \{0, 1, 2, \dots\}$. Then T (or the probability distribution of T) is said to have the *lack of memory property* if

$$P(\{T \geq j+k\}) \geq P(\{T \geq j\})P(\{T \geq k\}), \quad \forall j, k \in S_T. \blacksquare$$

The following theorem illustrates that, among all discrete probability distributions having the support $S = \{0, 1, 2, \dots\}$, the geometric distribution characterizes the lack of memory property.

Theorem 1.1

Let T be a discrete type random variable with support $S_T = \{0, 1, 2, \dots\}$. Then T has the lack of memory property if, and only if, $T \sim \mathrm{Ge}(p)$, for some $p \in (0,1)$.

Proof. We have seen that if $T \sim \mathrm{Ge}(p)$, for some $p \in (0,1)$, then T has the lack of memory property. Conversely suppose that T has the lack of memory property. Then

$$P(\{T \geq j+k\}) = P(\{T \geq j\})P(\{T \geq k\}), \ \forall j, k \in \{0,1,2, \dots\}.$$

Let $P(\{T = 0\}) = p$. Then $p \in (0,1)$, and, for $j \in \{0, 1, 2, , \dots\}$,

$$P(\{T \geq j+1\}) = P(\{T \geq j\})P(\{T \geq 1\})$$

$$= P(\{T \geq j\})(1-p)$$

$$= P(\{T \geq j-1\})(1-p)^2$$

$$\vdots$$

$$= P(\{T \geq 0\})(1-p)^{j+1}$$

$$= (1-p)^{j+1}.$$

Thus, for $k \in \{0, 1, 2, \dots\}$,

$$P(\{T = k\}) = P(\{T \geq k\}) - P(\{T \geq k+1\})$$

$$= (1-p)^k - (1-p)^{k+1}$$

$$= p(1-p)^k,$$

i.e., $T \sim \mathrm{Ge}(p)$. ■

Example 1.3

Consider a sequence of independent Bernoulli trials with probability of success in each trial being p. Let Z denote the number of trials required to get the r-th success, where r is a given positive integer. Let $X = Z - r$.

(i) Find the probability distributions of X and Z;

(ii) For $r = 1$, show that $P(\{Z > m+n\}) = P(\{Z > m\})P(\{Z > n\}), \forall m, n \in \{0, 1, \dots\}$(or equivalently $P(\{Z > m+n | Z > m\}) = P(\{Z > n\}), \forall m, n \in \{0, 1, \dots\}$; this property is also known as the lack of memory property).

Solution.

(i) Clearly $X = Z - r$ denotes the number of failures preceding the r-th success. Therefore $X \sim \mathrm{NB}(\mathrm{r}, p)$ and the p.m.f. of Z is given by

$$f_Z(z) = P(\{Z = z\})$$

$$= P(\{X = z - r\})$$

$$= \begin{cases} \binom{z-1}{r-1} p^r q^{z-r}, & \text{if } z \in \{r, r+1, \dots\} \\ 0, & \text{otherwise} \end{cases},$$

where $q = 1 - p$.

(ii) We have $X \sim \mathrm{NB}(\mathrm{r}, p)$ and $Z = X + 1$. Therefore, for $m, n \in \{0, 1, 2, \}$, $P(\{X \geq m+n\}) \geq P(\{X \geq m\})P(\{X \geq n\})$ and

$$P(\{Z > m+n\}) = P(\{X > m+n-1\})$$

$$= P(\{X \geq m+n\})$$
$$= P(\{X \geq m\})P(\{X \geq n\})$$
$$= P(\{Z \geq m+1\})P(\{Z \geq n+1\})$$
$$= P(\{Z > m\})P(\{Z > n\}). \blacksquare$$

Example 1.4

A person repeatedly rolls a fair dice independently until an upper face with two or three dots is observed twice. Find the probability that the person would require eights rolls to achieve this.

Solution. In each trial let us label the outcome of observing an upper face with two or three dots as success and observing any other outcome as a failure. Then we have a sequence of independent Bernoulli trials with probability of success in each trial as $p = \frac{1}{3}$. Let Z denote the number of trials required to get the second success. Then, using Example 1.3, $X \stackrel{\text{def}}{=} Z - 2 \sim \text{NB}(2, p)$. Therefore, the required probability is

$$P(\{Z = 8\}) = P(\{X = 6\}) = \binom{7}{1}\left(\frac{1}{3}\right)^2\left(1 - \frac{1}{3}\right)^6 = \frac{448}{6561}. \blacksquare$$

Example 1.5

Consider a person playing a sequence of games. Suppose that the games are played independently and the probability of person winning any game is $p \in (0, 1)$. Let the r.v. Z denote the number of games the person will have to play to record the first win. Then, by Example 1.3, $P(\{Z > 12\}|\{Z > 10\}) = P(\{Z > 2\})$, i.e., the conditional probability that the person will require at least three additional games to record the first win, given that the person has lost the first ten games, is the same as the probability that he will require at least three games, since he started playing, to record the first win. Here the information ($\{Z > 10\}$) that the person has lost first ten games has no bearing on the additional number of games he will require to record a win. $\blacksquare$

Example 1.6

Two teams (say Team A and Team B) play a series of games until one team wins 5 games. If the probability of Team A (Team B) winning any game is 0.7 (0.3), find the probability that the series will end in 8 games.

Solution. Let X_1 (X_2) denote the number of games Team A (Team B) will have to play to secure the fifth win. Then the required probability is $p = P(\{X_1 = 8\}) + P(\{X_2 = 8\})$. By Example 1.3, we have $X_1 = Y_1 + 5$ and $X_2 = Y_2 + 5$, where $Y_1 \sim \text{NB}(5, 0.7)$ and $Y_2 \sim \text{NB}(5, 0.3)$. Therefore

$$P(\{X_1 = 8\}) = P(\{Y_1 = 3\}) = \binom{7}{4}(0.7)^5(1-0.7)^3 = .1588 \text{ (approximately)},$$

$$P(\{X_2 = 8\}) = P(\{Y_2 = 3\}) = \binom{7}{4}(0.3)^5(1-0.3)^3 = .0292 \text{ (approximately)}$$

and the required probability is

$$p = P(\{X_1 = 8\}) + P(\{X_2 = 8\}) = 0.188 \text{ (approximately)}. \blacksquare$$

4.3 THE HYPERGEOMETRIC DISTRIBUTION

Consider a population comprising of $N\ (\geq 2)$ units out of which $a\ (\in \{1,2,\dots,N-1\})$ are labeled as S (success) and $N-a$ are labeled as F (failure). Suppose that we are interested in drawing a sample of size $n\ (\in \{1,2,\dots,N-1\})$ from this population, drawing one unit at a time. Let X denote the number of successes (S) in the drawn sample. We consider the following two cases.

Case I: Draws are independent and sampling is with replacement (i.e., after each draw the drawn unit is replaced back into the population)

A sampling of this kind is called a *simple random sampling with replacement*. As discussed in Remark 1.1, in this case, we have a sequence of n independent Bernoulli trials with probability of success is each trial as $p = \frac{a}{N}$ and, therefore, $X \sim \text{Bin}\left(n, \frac{a}{N}\right)$.

Case II: Sampling is without replacement (i.e., drawn units are not replaced back into the population)

A sampling of this kind is called a *simple random sampling without replacement*. In this case

$$P(\{\text{obtaining } S \text{ in first draw}\}) = \frac{a}{N};$$

$$P(\{\text{obtaining } S \text{ in second draw}\}) = \frac{a}{N} \cdot \frac{a-1}{N-1} + \frac{N-a}{N} \cdot \frac{a}{N-1} = \frac{a}{N};$$

$$P(\{\text{obtaining } S \text{ in third draw}\}) = \frac{a}{N} \cdot \frac{a-1}{N-1} \cdot \frac{a-2}{N-2} + \frac{a}{N} \cdot \frac{N-a}{N-1} \cdot \frac{a-1}{N-2}$$
$$+ \frac{N-a}{N} \cdot \frac{a}{N-1} \cdot \frac{a-1}{N-2} + \frac{N-a}{N} \cdot \frac{N-a-1}{N-1} \cdot \frac{a}{N-1}$$

$$= \frac{a}{N}.$$

In general, it can be shown that (see Theorem 2.1 in the sequel)

$$P(\{\text{obtaining } S \text{ in } k-\text{th draw}\}) = \frac{a}{N}\ , \quad k = 1,2, \dots, n.$$

Note that

$$P(\{\text{obtaining } S \text{ in first two draws}\}) = \frac{a}{N} \cdot \frac{a-1}{N-1}$$

$$\neq \frac{a}{N} \cdot \frac{a}{N}$$

$$= P(\{\text{obtaining } S \text{ in first draw}\}) \times P(\{\text{obtaining } S \text{ in second draw}\})\ ,$$

implying that the draws are not independent. Thus, in this case, although we have a sequence of Bernoulli trials with the same probability ($p = \frac{a}{N}$) of success in each trial, the trials are not independent. Therefore, in this case, we cannot conclude that $X \sim \text{Bin}(n, \frac{a}{N})$.

It can be seen that the distribution of X remains the same whether we sample one by one without replacement or select a subset of size n at random (so that each of $\binom{N}{n}$ subsets of size n has the same probability of getting selected as desired sample) . Clearly, for $P(\{X = k\})$ to be non-zero, we must have $k \in \mathbb{N},\ 0 \le k \le n,\ 0 \le k \le a$ and $0 \le n - k \le N - a$, i.e., $k \in S_X = \{m \in \mathbb{N}: \max(0, n - N + a) \le m \le \min(n, a)\}$. Note that, for $x \in S_X$, the event $\{X = x\}$ occurs if, and only if, the selected subset (sample) contains x successes and $n - x$ failures. It follows that the total number of equally likely cases favorable to the event $\{X = x\}$ is $\binom{a}{x}\binom{N-a}{n-x}$. Since the total number of possible subsets (samples) of size n of N units is $\binom{N}{n}$, the p.m.f. of X is given by

$$f_X(x) = \begin{cases} \dfrac{\binom{a}{x}\binom{N-a}{n-x}}{\binom{N}{n}}, & \text{if } x \in S_X = \{\max(0, n-N+a), \dots, \min(n,a)\} \\ 0, & \text{otherwise} \end{cases}. \qquad (2.1)$$

The probability distribution with p.m.f. (2.1) is called the *Hypergeometric distribution* and is denoted by $\text{Hyp}(a, n, N), a, n, N \in \mathbb{N}, N \ge 2, a, n \le N - 1$. We shall use the notation $X \sim \text{Hyp}(a, n, N)$ to indicate that the r.v. X follows $\text{Hyp}(a, n, N)$ distribution. Clearly we have a family $\{\text{Hyp}(a, n, N), a, n, N \in \mathbb{N}, N \ge 2, a, n \le N - 1\}$ of hypergeometric distributions corresponding to different choices of (a, n, N) such that $N \in \{2, 3, \dots\}, a \in \{1, 2, \dots, N-1\}$ and $n \in \{1, 2, \dots, N-1\}$.

Since $\sum_{x \in S_X} f_X(x) = 1$, we have the following identity

$$\sum_{k=\max(0, n-N+a)}^{\min(n,a)} \binom{a}{k}\binom{N-a}{n-k} = \binom{N}{n}. \qquad (2.2)$$

For a positive integer r, the r-th factorial moment of X is given by

$$E(X_{(r)}) = E\left(\prod_{j=0}^{r-1}(X-j)\right)$$

$$= \frac{1}{\binom{N}{n}} \sum_{k=\max(0, n-N+a)}^{\min(n,a)} \left\{\prod_{j=0}^{r-1}(k-j)\right\} \binom{a}{k}\binom{N-a}{n-k}.$$

Clearly, for $r \in \mathbb{N}$ and $r > \min(n, a)$, $E(X_{(r)}) = 0$. For $r \in \mathbb{N}$ and $r \leq \min(n, a)$, we have

$$E(X_{(r)}) = \frac{1}{\binom{N}{n}} \sum_{k=\max(r, n-N+a)}^{\min(n,a)} \left\{\prod_{j=0}^{r-1}(k-j)\right\} \binom{a}{k}\binom{N-a}{n-k}$$

$$= \frac{a_{(r)}}{\binom{N}{n}} \sum_{k=\max(r, n-N+a)}^{\min(n,a)} \binom{a-r}{k-r}\binom{N-a}{n-k}$$

$$= \frac{a_{(r)}}{\binom{N}{n}} \sum_{k=\max(0, n-N+a-r)}^{\min(n-r, a-r)} \binom{a-r}{k}\binom{N-a}{n-\mathrm{r}-k}$$

$$= \frac{a_{(r)}}{\binom{N}{n}} \sum_{k=\max(0, (n-r)-(N-r)+a-r)}^{\min(n-r, a-r)} \binom{a-r}{k}\binom{(N-r)-(a-r)}{(n-r)-k}$$

$$= \frac{\binom{N-r}{n-r}}{\binom{N}{n}} a_{(r),} \qquad \text{(using (2.2))}$$

where $a_{(r)} = a(a-1)\cdots(a-r+1)$.

Thus, for $r \in \mathbb{N}$, we have

$$E\big(X_{(r)}\big) = \begin{cases} \dfrac{\binom{N-r}{n-r}}{\binom{N}{n}} a_{(r)}, & \text{if } r \le \min(n,a) \\ 0, & \text{if } r > \min(n,a) \end{cases}.$$

In particular

$$E(X) = E\big(X_{(1)}\big) = n\frac{a}{N} = np,$$

where $p = a/N$. For $n \ge 2, a \ge 2$

$$E\big(X(X-1)\big) = E\big(X_{(2)}\big) = n(n-1)\frac{a(a-1)}{N(N-1)};$$

$$\begin{aligned} \mathrm{Var}(X) &= E(X^2) - \big(E(X)\big)^2 \\ &= E\big(X(\mathrm{X}-1)\big) + E(X) - \big(E(X)\big)^2 \\ &= n\left(\frac{a}{N}\right)\left(1-\frac{a}{N}\right)\frac{N-n}{N-1} \end{aligned}$$

$$\mathrm{Var}(X) = np(1-p)\left(1-\frac{n-1}{N-1}\right). \qquad (2.3)$$

Remark 2.1

In the case of sampling with replacement we have $X \sim \mathrm{Bin}\left(n,\frac{a}{N}\right), E(X) = np = n\frac{a}{N}$ and $\mathrm{Var}(X) = np(1-p) = n\frac{a}{N}\left(1-\frac{a}{N}\right)$, where $p = \frac{a}{N}$. The factor $(1-\frac{n-1}{N-1})$, which on multiplying to the variance of $\mathrm{Bin}\left(n,\frac{a}{N}\right)$ distribution yields the variance of $\mathrm{Hyp}(a,n,N)$ distribution (see (2.3)), is called the *finite population correction* (f.p.c.). Clearly if the sample size n is significantly smaller than the population size N ($n \ll N$) then the f.p.c. will be close to 1 and therefore the variances of $\mathrm{Bin}\left(n,\frac{a}{N}\right)$ and $\mathrm{Hyp}(a,n,N)$ distributions will be very close. In fact we will see later (Theorem 2.2) that when $n \ll N$ and $n \ll a \equiv a_N$ (say) are such that $\frac{a_N}{N}$ is a fixed quantity (i.e., as $N \to \infty, a_N \to \infty$ and $\frac{a_N}{N} \to p$, where $p \in (0,1)$ is a fixed quantity) then $\mathrm{Bin}\left(n,\frac{a}{N}\right)$ provides an approximation to $\mathrm{Hyp}(a,n,N)$

distribution (see Figures 2.1 and 2.2). Regarding choice of sample size n for using this approximation, a guideline, based on various empirical studies, is that the sample size n should not exceed 10% of the population size N. ■

Since the support $S_X = \{m \in \mathbb{N}: \max(0, n - N + a) \leq m \leq \min(n, a)\}$ is finite the m.g.f. $M_X(t) = E(e^{tX})$ exists, although a closed form expression for it cannot be obtained.

Theorem 2.1

Under the above notation, let $A_i, i = 1, 2, \dots, n$, denote the probability of observing a success in the i-th trial. Then $P(A_i) = \frac{a}{N}, i = 1, 2, \dots, n$, i.e., the probability of success in each trial is the same.

Proof. Clearly $P(A_1) = \frac{a}{N}$ and

$$\begin{aligned} P(A_2) &= P(A_1^C \cap A_2) + P(A_1 \cap A_2) \\ &= P(A_1^C)P(A_2|A_1^C) + P(A_1)P(A_2|A_1) \\ &= \frac{N-a}{N} \cdot \frac{a}{N-1} + \frac{a}{N} \cdot \frac{a-1}{N-1} \\ &= \frac{a}{N}. \end{aligned}$$

Now suppose that $P(A_m) = \frac{a}{N}$, for some $m \in \{1, 2, \dots, n-1\}$. Let us denote the probability mass function defined in (2.1) by $p(x|a, n, N), x \in \mathbb{N}$ and the corresponding r.v. by $X_{a,n,N}$. Then the r.v. $X_{a,m,N}$ denotes the number of successes in the first m trails and, therefore,

$$\begin{aligned} P(A_{m+1}) &= \sum_{k=0}^{m} P(\{X_{a,m,N} = k\}) P(A_{m+1} | \{X_{a,m,N} = k\}) \\ &= \sum_{k=0}^{m} p(k|a, m, N) \frac{a-k}{N-m} \\ &= \sum_{k=\max(0, m-N+a)}^{\min(m,a)} \frac{\binom{a}{k}\binom{N-a}{m-k}}{\binom{N}{m}} \frac{a-k}{N-m} \end{aligned}$$

$$= \frac{a}{N-m} \sum_{k=\max(0,m-N+a)}^{\min(m,a)} \frac{\binom{a}{k}\binom{N-a}{m-k}}{\binom{N}{m}} - \frac{1}{N-m} \sum_{k=\max(0,m-N+a)}^{\min(m,a)} k \frac{\binom{a}{k}\binom{N-a}{m-k}}{\binom{N}{m}}$$

$$= \frac{a}{N-m} - \frac{1}{N-m} E(X_{a,m,N})$$

$$= \frac{a}{N-m} - \frac{1}{N-m} m \frac{a}{N}$$

$$= \frac{a}{N}.$$

The result now follows by principle of mathematical induction. ■

Remark 2.2

Under the hypergeometric distribution described above we have a sequence of n dependent Bernoulli trials with probability of success in each trial being $p = a/N$. The hypergeometric distribution provides the distribution of X, the number of successes in $n \in \{1, 2, \ldots, N-1\}$ dependent Bernoulli trials. If the trials were independent the distribution of X would be provided by the binomial distribution. Our intuition suggests that, under sampling without replacement from an infinite population having infinite number of successes, the Bernoulli trials should be independent. This is in fact true and it can be shown that if $N \to \infty, a \equiv a_N \to \infty$ and $\frac{a_N}{N} \to p$ (where $p \in (0,1)$ is a fixed quantity) then the trials are independent. The above discussion suggests that if the population size N and a (the number of S in the population) are infinite and $\frac{a}{N} = p \in (0,1)$ is a fixed quantity then the distribution of X remains the same whether the sampling is done with replacement or without replacement. In such situations, therefore, one may use either hypergeometric or binomial distribution to provide the distribution of X. The following theorem corroborates this observation. ■

Theorem 2.2

Let $X_{a_N,n,N} \sim \text{Hyp}(a_N, n, N)$, where a_N depends on N and $\lim_{N\to\infty} \frac{a_N}{N} = p \in (0,1)$. Let $f_{a_N,n,N}(\cdot)$ denote the p.m.f. of $X_{a_N,n,N}$. Then

$$\lim_{N\to\infty} f_{a_N,n,N}(k) = \lim_{N\to\infty} P(\{X_{a_N,n,N} = k\}) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k}, & \text{if } k \in \{0,1,\ldots,n\} \\ 0, & \text{otherwise} \end{cases}.$$

i.e., for large N and large a_N, so that $p = \frac{a_N}{N}$ is a fixed quantity, $\text{Hyp}(a_N, n, N)$ probabilities can be approximated by $\text{Bin}\left(n, \frac{a}{N}\right)$ probabilities.

Proof. For notational convenience let us denote $X_{a_N,n,N}$ by X. The support of r.v. X is $S_X = \{m \in \mathbb{N}: \max(0, n-N+a_N) \le m \le \min(n, a_N)\}$. Note that

$$n - N + a_N = N\left(\frac{n}{N} - 1 + \frac{a_N}{N}\right) \to -\infty, \text{ and } \quad a_N = N\frac{a_N}{N} \to \infty, \text{ as } N \to \infty.$$

It follows that, as $N \to \infty$,

$$\max(0, n-N+a_N) \to 0 \quad \text{and } \min(n, a_N) \to n.$$

Also, for $k \in S_X$,

$$f_X(k) = \frac{\binom{a_N}{k}\binom{N-a_N}{n-k}}{\binom{N}{n}}$$

$$= \binom{n}{k}\left\{\prod_{j=0}^{k-1}\left(\frac{a_N - j}{N-j}\right)\right\}\left\{\prod_{j=0}^{n-k-1}\left(\frac{N-a_N-j}{N-k-j}\right)\right\}$$

$$\xrightarrow{N\to\infty} \binom{n}{k}\left\{\prod_{j=0}^{k-1}(p)\right\}\left\{\prod_{j=0}^{n-k-1}(1-p)\right\}$$

$$= \binom{n}{k} p^k (1-p)^{n-k}.$$

Therefore

$$\lim_{N\to\infty} f_{a_N,n,N}(k) = \begin{cases}\binom{n}{k} p^k (1-p)^{n-k}, & \text{if } k \in \{0, 1, \dots, n\} \\ 0, & \text{otherwise}\end{cases}. \blacksquare$$

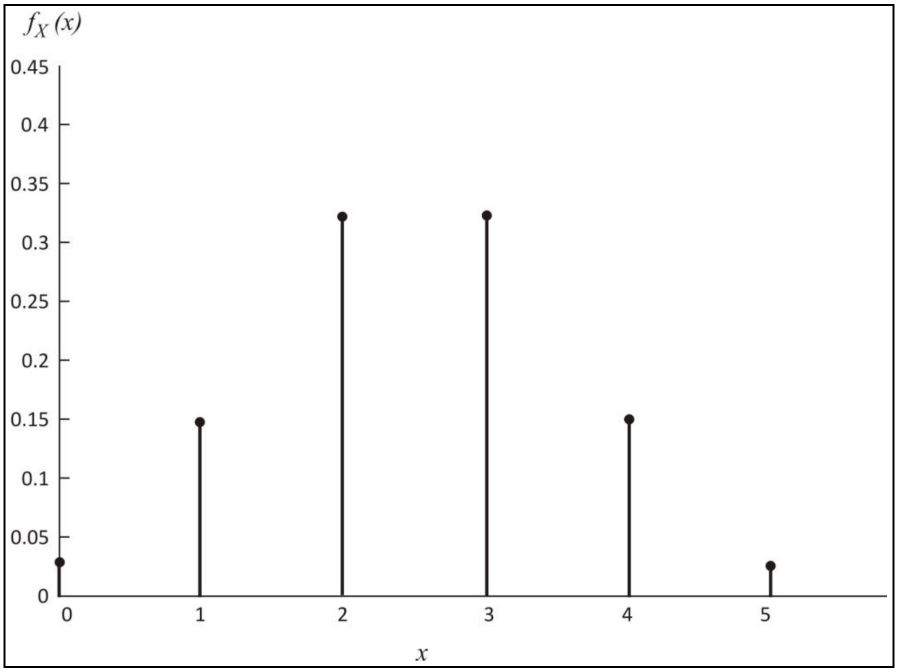

Figure 2.1. Plot of p.m.f. of Hyp(25, 5,50)

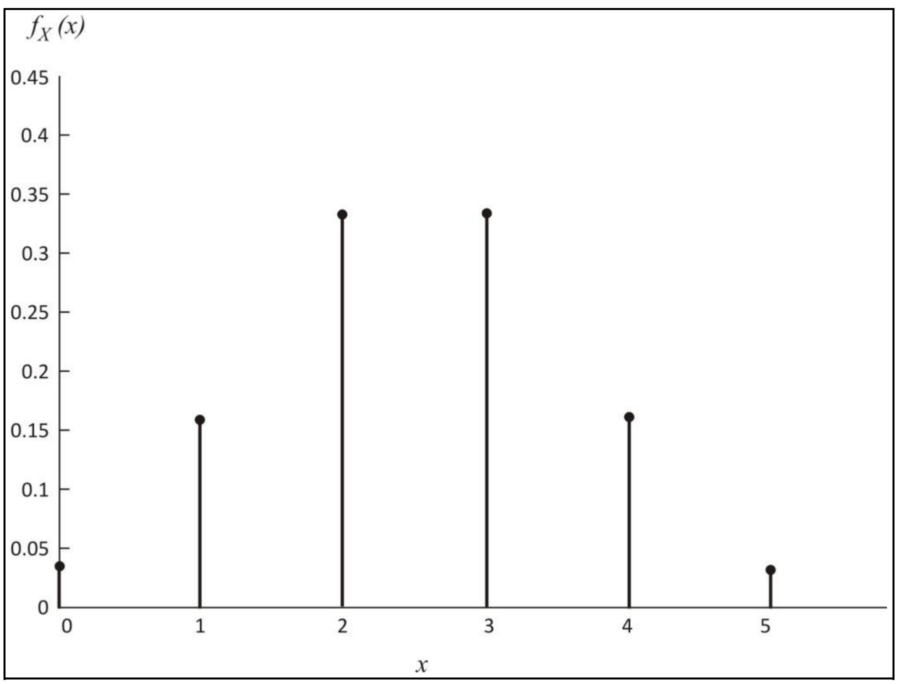

Figure 2.2. Plot of p.m.f. of Bin$(5, \frac{1}{2})$