Chapter IV Discrete Distributions

The probability models for random experiments that will be described in this and next chapters occur frequently in applications. Continuous distributions will be presented in next chapter. This chapter will introduce some discrete distributions, including Bernoulli distribution, binomial distribution, Poisson distribution, geometric distribution, negative binomial distribution, and hypergeometric distribution.

4.1 Bernoulli Trials and Bernoulli Distributions

On a single trial of an experiment, suppose that there are only ***two events*** of interest, say E and E^c. For example, E and E^c could represent the occurrence of a "head" or a "tail" on a single coin toss, obtaining a "defective" or a "good" item when drawing a single item from a manufactured lot, or, in general, "success" or "failure" on a particular trial of an experiment. Suppose that E occurs with probability $p = P(E)$, and consequently E^c occurs with probability $q = P(E^c) = 1 - p$.

A random variable, *X*, that assumes only the value 0 or 1 is known as a ***Bernoulli variable***, and a performance of an experiment with only two types of outcomes is called a ***Bernoulli trial***. In particular, if an experiment can result only in "success" (E) or "failure" (E^c), then the corresponding Bernoulli variable is

$$X(e) = \begin{cases} 1 & \text{if } e \in E \\ 0 & \text{if } e \in E^c. \end{cases}$$

The pmf of *X* is given by $f(0) = q(= 1 - p)$ and $f(1) = p$. The corresponding distribution is known as a ***Bernoulli distribution*** with the ***parameter*** *p*, and its pmf can be expressed as

$$f(x) = p^x q^{1-x} \qquad x = 0, 1.$$

It is obvious that the probabilities sum to one, $\sum_{x=0}^{1} f(X) = f(0) + f(1) = p^0 q^1 + p^1 q^0 = 1$.

Theorem 4.1-1: If *X* is a random variable with a Bernoulli distribution, then

$$\mu = p, \quad \sigma^2 = pq, \text{ and } \quad M(t) = pe^t + q \qquad -\infty < t < \infty$$

Proof: $\mu = E(X) = \sum_{x=0}^{1} xf(x) = 0f(0) + 1f(1) = 0 \times q + 1 \times p = p$.

$$E(x^2) = \sum_{x=0}^{1} x^2 f(x) = 0^2 f(0) + 1^2 f(1) = 0^2 \times q + 1^2 \times p = p.$$

$$\sigma^2 = E(X^2) - [E(X)]^2 = p - p^2 = p(1-p) = pq.$$

$$M(t) = E(e^{tX}) = \sum_{x=0}^{1} e^{tx} f(x) = e^{t\times 0} f(0) + e^{t\times 1} f(1) = q + pe^t \qquad -\infty < t < \infty. \quad \blacksquare$$

4.2 Binomial Distribution

Often it is possible to structure a more complicated experiment as a sequence of ***independent Bernoulli trials***, where the quantity of interest is the number of successes on a ***certain numbers*** of trials. In a sequence of ***n*** *independent Bernoulli trials* with probability of **success** p and probability of ***failure*** $q = (1 - p)$ on each trial, let X represent the number of successes. The discrete pmf of X, known as the binomial distribution and denoted by $X \sim B(n, p)$, is given by

$$f(x) = \binom{n}{x} p^x q^{n-x} \qquad x = 0, 1, 2, \cdots, n.$$

The quantities of n and p are called the ***parameters of the binomial distribution***. **Note** that by the binomial theorem, the probabilities sum to one; that is

$$\sum_{x=0}^{n} f(X) = \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x} = (p+q)^n = 1^n = 1$$

Example 4.2-1: It is known that screws produced by a certain company will be defective with probability 0.01 independently of each other. The company sells the screws in package of 10 and offers a money-back guarantee that at most 1 of the 10 screws is defective. What proportion of packages sold must the company replace?

Solution: If X is the number of defective screws in a package, then X is a binomial random variable with parameters (10, 0.01). Hence, the probability that a package will have to be replaced is $1 - P(X = 0) - P(X = 1) = 1 - \binom{10}{0}(.01)^0(.99)^{10} - \binom{10}{1}(.01)^1(.99)^9 \approx 0.004$.

Hence, only 0.4 percent of the package will have to be replaced. ◘

Summarizing, a binomial experiment satisfies the following properties:

1. A Bernoulli (success-failure) experiment is performed ***n* times**.
2. The trials are ***independent***.
3. The probability of success on each trial is a ***constant*** p; the probability of failure is $q = 1 - p$.
4. The random variable X counts the ***number of successes*** in the n trials.

Theorem 4.2-1: If X is a random variable with a binomial distribution with parameter n and p, $X \sim B(n, p)$, then

$$\mu = np, \quad \sigma^2 = npq, \text{ and } \quad M(t) = (pe^t + q)^n \qquad -\infty < t < \infty$$

Proof: $E(X) = \sum_{x=0}^{n} x f(X) = \sum_{x=0}^{n} x \binom{n}{x} p^x q^{n-x} = \sum_{x=0}^{n} x \frac{n!}{x!(n-x)!} p^x q^{n-x} = \sum_{x=1}^{n} \frac{n!}{(x-1)!(n-x)!} p^x q^{n-x}$

$$= np \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} q^{n-x} \underbrace{=}_{\text{let } y = x-1} np \sum_{y=0}^{n-1} \frac{(n-1)!}{y!(n-1-y)!} p^y q^{n-1-y}$$

$$= np(p+q)^{n-1} = np.$$

$$E[X(X-1)] = \sum_{x=0}^{n} x(x-1)\binom{n}{x} p^x q^{n-x} = \sum_{x=0}^{n} x(x-1)\frac{n!}{x!(n-x)!} p^x q^{n-x}$$

$$= \sum_{x=2}^{n} \frac{n!}{(x-2)!(n-x)!} p^x q^{n-x} = n(n-1)p^2 \sum_{x=2}^{n} \frac{(n-2)!}{(x-2)!(n-x)!} p^{x-2} q^{n-x}$$

$$\underbrace{=}_{\text{let } y=x-2} n(n-1)p^2 \sum_{y=0}^{n-2} \frac{(n-2)!}{y!(n-2-y)!} p^y q^{n-2-y}$$

$$= n(n-1)p^2 (p+q)^{n-2} = n(n-1)p^2.$$

$$\sigma^2 = E(X^2) - [E(X)]^2 = [E(X(X-1)) + E(X)] - [E(X)]^2 = [n(n-1)p^2 + np] - (np)^2$$

$$= -np^2 + np = np(1-p) = npq.$$

$$M(t) = E(e^{tX}) = \sum_{x=0}^{n} e^{tx}\binom{n}{x} p^x q^{n-x} = \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x q^{n-x} = (pe^t + q)^n \qquad -\infty < t < \infty. \quad \square$$

Example 4.2-2: A communication system consists of n components each of which will, independently, function with probability p. The total system will be able to operate effectively if at least one-half of its components function. For what value of p is a 5-component system more likely to operate effectively than 3-component system?

Solution: As the number of functioning components is a binomial random variable with parameters (n, p), it follows that the probability that a 5-component system will be effective is

$$\binom{5}{3} p^3 (1-p)^2 + \binom{5}{4} p^4 (1-p)^1 + \binom{5}{5} p^5$$

whereas the corresponding probability for a 3-component system is

$$\binom{3}{2} p^2 (1-p)^1 + \binom{3}{3} p^3.$$

Hence, the 5-component system is better if

$$10p^3(1-p)^2 + 5p^4(1-p) + p^5 \geq 3p^2(1-p) + p^3$$

which reduces to

$$3(p-1)^2(2p-1) \geq 0$$

or $$p \geq 1/2. \quad \square$$

4.3 Geometric Distribution and Negative Binomial Distribution

We turn now to the problem of observing a sequence of Bernoulli trials until **exactly *r*** successes occur, where *r* is a fixed positive integer. Let the random variable *X* denote the number of trials needed to observe the *r*th success. That is, *X* is the trial number on which the *r*th success is observed. By the multiplication rule of probabilities, the pmf of *X*, say $f(x)$, equals the product of the probability

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

of obtaining exactly $r-1$ successes in the first $x-1$ trials and the probability *p* of a success on the *r* trial. Thus the pmf of *X* is

$$f(x) = \binom{x-1}{r-1}p^{r}(1-p)^{x-r} = \binom{x-1}{r-1}p^{r}q^{x-r} \qquad x = r, r+1, \cdots .$$

We say that *X* has a **negative binomial distribution,** denoted by $X \sim \text{NB}(r, p)$ with the ***parameters*** *r* and *p*.

We first discuss this problem when $r = 1$. That is, consider a sequence of Bernoulli trials with probability *p* of success. This sequence is observed until the first success occurs. We say that *X* has a **geometric** (**Pascal**) **distribution,** denoted by $X \sim \text{GEO}(p)$**,** with the ***parameter*** *p* since the pmf consists of terms of a geometric series, namely

$$f(x) = p(1-p)^{x-1} \qquad x = 1, 2, \cdots .$$

It can be verified that the probabilities sum to one, $\sum_{x=1}^{\infty} p(1-p)^{x-1} = \dfrac{p}{1-(1-p)} = 1$.

Example 4.3-1: Suppose that the probability of engine malfunction during any 1-hour period is $p = 0.02$. Find the probability that a given engine will survive 2 hours.

Solution: Letting *X* denote the number of 1-hour intervals until until the first malfunction, we have

$$P(\text{survive 2 hours}) = P(X \geq 3) = \sum_{x=3}^{\infty} pq^{x-1} = 1 - \sum_{x=1}^{2} pq^{x-1} = 1 - p - pq = 0.9604 . \ \square$$

Theorem 4.3-1: If *X* is a random variable with a geometric distribution with parameter *p*, then

$$\mu = 1/p , \qquad \sigma^2 = q/p^2 , \qquad \text{and} \qquad M(t) = \frac{pe^t}{1-qe^t} , \qquad qe^t < 1$$

Proof: To find the mean and the variance for the geometric distribution, we will use the following results about the sum and the first and second derivatives of a geometric series. For $-1 < r < 1$, let

$$g(r) = \sum_{k=0}^{\infty} ar^k = \frac{a}{1-r} .$$

Then, $$g'(r) = \sum_{k=1}^{\infty} akr^{k-1} = \frac{a}{(1-r)^2}$$

and $$g''(r) = \sum_{k=2}^{\infty} ak(k-1)r^{k-2} = \frac{2a}{(1-r)^3}.$$

$$E(X) = \sum_{x=1}^{\infty} xf(x) = \sum_{x=1}^{\infty} xq^{x-1}p = \frac{p}{(1-q)^2} = \frac{1}{p}.$$

$$E(X(X-1)) = \sum_{x=1}^{\infty} x(x-1)q^{x-1}p = \sum_{x=2}^{\infty} pqx(x-1)q^{x-2} = \frac{2pq}{(1-q)^3} = \frac{2q}{p^2}.$$

$$\sigma^2 = E[X(X-X)] + E(X) - [E(x)]^2 = \frac{2q}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{q}{p^2}.$$

$$M(t) = \sum_{x=1}^{\infty} e^{tx} pq^{x-1} = pe^t \sum_{x=1}^{\infty} \left(qe^t\right)^{x-1} = \frac{pe^t}{1-qe^t} \qquad |\, qe^t \,| < 1. \quad \square$$

Theorem 4.3-2: $P(X > k) = q^k$.

Proof: $P(X > k) = \sum_{x=k+1}^{\infty} pq^{x-1} = \frac{pq^{(k+1)-1}}{1-q} = q^k$. $\square$

Theorem 4.3-3: No-Memory Property

If $X \sim \text{GEO}(p)$**,** then $P(X > j + k \mid X > j) = P(X > k)$ where *j* and *k* are nonnegative integers.

Proof: Left as exercise.

Example 4.3-2: Consider the problem of obtaining a random ordering of the first *n* positive integers. For example, suppose that we would like a random ordering of the first positive integers. We would obtain this random ordering by rolling a ***fair*** six-sided die. The first cast of the die would give the first outcome in the random ordering. To obtain the second number in the ordering, only five of the six possible outcomes are eligible. After the first $k-1$ positions have been filled with unique integers, the number of candidates for position *k* is $6-k+1$ for *k* = 1, 2, 3, 4, 5, 6. The probability of selecting one of these eligible integers is $p_k = (6-k+1)/6$. If X_k denotes the number of trials needed to observe the first success (an integer that has not yet been selected), then X_k has a geometric distribution with $p = p_k$.

An interesting problem is to determine the average number of casts of the die to obtain a random ordering of 1, 2, 3, 4, 5, 6. Let X_k equal the number of casts required to fill position *k*. If we let $W = X_1 + X_2 + X_3 + X_4 + X_5 + X_6$, then W denotes the total number of casts required. So the average number of casts required is

$$E(W) = E(X_1) + E(X_2) + \cdots + E(X_6) = 1 + \frac{1}{5/6} + \frac{1}{4/6} + \cdots + \frac{1}{1/6} = 14.7\,. \quad \square$$

Let us now turn to the negative binomial distribution. The reason for calling this the negative binomial distribution is the following. Consider $h(w) = (1-w)^{-r}$, the binomial $(1-w)$ with the negative exponent $-r$. Using Maclaurin's series expansion, we have

$$(1-w)^{-r} = \sum_{k=0}^{\infty} \frac{h^{(k)}(0)}{k!} w^k = \sum_{k=0}^{\infty} \binom{r+k-1}{r-1} w^k\,.$$

If we let $x = k + r$ in the summation, then $k = x - r$ and

$$(1-w)^{-r} = \sum_{x=r}^{\infty} \binom{r+x-r-1}{r-1} w^{x-r} = \sum_{x=r}^{\infty} \binom{x-1}{r-1} w^{x-r},$$

the summand of which is, except for the factor p^r, the negative binomial probability when $w = q$.

According the above formula, we obtain

$$\sum_{x=r}^{\infty} f(x) = \sum_{x=r}^{\infty} \binom{x-1}{r-1} p^r (1-p)^{x-r} = p^r [1-(1-p)]^{-r} = 1$$

Theorem 4.3-4: If X is a random variable with a negative binomial distribution with parameters r and p, $X \sim \mathrm{NB}(r, p)$, then

$$\mu = r/p\,, \qquad \sigma^2 = rq/p^2\,, \qquad \text{and} \qquad M(t) = \frac{(pe^t)^r}{(1-qe^t)^r}, \qquad |qe^t| < 1\,.$$

Proof: To find the mean and variance, we will use the following derivatives of the series expansion of the negative binomial. Let

$$h(w) = \sum_{k=0}^{\infty} \binom{r+k-1}{r-1} w^k = (1-w)^{-r}\,.$$

Then

$$h'(w) = \sum_{k=1}^{\infty} \binom{r+k-1}{r-1} k w^{k-1} = r(1-w)^{-r-1}$$

and

$$h''(w) = \sum_{k=2}^{\infty} \binom{r+k-1}{r-1} k(k-1) w^{k-2} = r(r+1)(1-w)^{-r-2}\,.$$

$$E(X-r)=\sum_{x=r}^{\infty}(x-r)\binom{x-1}{r-1}p^r q^{x-r}=\sum_{x=r+1}^{\infty}(x-r)\binom{x-1}{r-1}p^r q^{x-r}\,.$$

$$\underbrace{=}_{\text{Let } k=x-r}\ \sum_{k=1}^{\infty}k\binom{r+k-1}{r-1}p^r q^k = p^r q\sum_{k=1}^{\infty}\binom{r+k-1}{r-1}kq^{k-1}$$

$$= p^r qr(1-q)^{-r-1} = rq/p\,.$$

$$E(X) = r + rq/p = r/p\,.$$

$$E[(X-r)(X-r-1)]=\sum_{x=r}^{\infty}(x-r)(x-r-1)\binom{x-1}{r-1}p^r q^{x-r}$$

$$= p^r q^2\sum_{x=r+2}^{\infty}\binom{x-1}{r-1}(x-r)(x-r-1)q^{x-r-2}$$

$$\underbrace{=}_{\text{Let } k=x-r}\ p^r q^2\sum_{k=2}^{\infty}\binom{k+r-1}{r-1}k(k-1)q^{k-2} = p^r q^2 r(r+1)(1-q)^{-r-2} = \frac{q^2}{p^2}r(r+1)\,.$$

$$\sigma^2 = \mathrm{Var}(X) = \mathrm{Var}(X-r) = E[(X-r)(X-r-1)] + E(X-r) - [E(X-r)]^2$$

$$= \frac{q^2}{p^2}r(r+1) + r\frac{q}{p} - r^2\frac{q^2}{p^2} = \frac{rq}{p^2}\,.$$

$$M(t) = \sum_{x=r}^{\infty}e^{tx}\binom{x-1}{r-1}p^r q^{x-r} = \left(pe^t\right)^r\sum_{x=r}^{\infty}\binom{x-1}{r-1}\left(qe^t\right)^{x-r} = \frac{\left(pe^t\right)^r}{\left(1-qe^t\right)^r}\qquad |\,qe^t\,|<1\,.\quad ◘$$

Example 4.3-3: Team A plays team B in a seven-game series. That is, the series is over when either team wins four games. For each game, $P(A \text{ wins}) = 0.6$, and the games are assumed independent. What is the probability that the series will end in exactly six games?

Solution: Let X be the number of games in the series. Then $X \sim \mathrm{NB}(4, 0.6)$ if team A wins the series and $X \sim \mathrm{NB}(4, 0.4)$ if team B wins the series. Hence, we have

$$P(A\text{ wins series in 6}) = \binom{5}{3}(0.6)^4(0.4)^2 = 0.20736$$

$$P(B\text{ wins series in 6}) = \binom{5}{3}(0.4)^4(0.6)^2 = 0.09216$$

$$P(\text{series goes 6 games}) = 0.20736 + 0.09216 = 0.29952\,.\quad ◘$$

The negative binomial problem is sometimes referred to as **inverse binomial sampling**. Suppose that $X \sim \mathrm{NB}(r, p)$ and $W \sim \mathrm{B}(n, p)$. It follows that

$$P(X \le n) = P(W \ge r)\,.$$

That is, $W \geq r$ corresponding to the event of having r or more successes in n trials, and that means n or fewer trials will be needed to obtain the first r successes.

Note that the number of experiments in the binomial experiment is a fixed number, while it is a random variable in both the geometric and the negative binomial distributions. We summarize the characteristics of these three distributions in Table 4-1.

Table 4-1

	Binomial Dist.	**Geometric** Dist.	**Negative Binomial** Dist.
Type of Experiments	Independent Bernoulli Trials with the Probability of Success p		
Number of **Successes**	**Random** Variable	**Fixed** Number (1)	**Fixed** Number (r)
Number of **Experiments**	**Fixed** Number (n)	**Random** Variable	**Random** Variable

4.4 Poisson Distribution

Some experiments result in counting the number of times particular events occur in given times or on given physical objects. For example, we could count the number of phone calls arriving at a switchboard between 9 and 10 A.M., the number of flaws in 100 feet of wire, the number of customers that arrive at a ticket window between 12 noon and 2 P.M., or the number of defects in a 100-foot roll of aluminum screen that is 2 feet wide. Each count can be looked upon as a random variable associated with an approximate Poisson process provided the condition in Definition 4-1 are satisfied.

Definition 4.4-1: Let the number of changes that occur in a given continuous interval be counted. We have an ***approximate Poisson process*** with parameter $\lambda > 0$ if the following are satisfied:

(i) The numbers of changes occurring in **non-overlapping** intervals are **independent**.

(ii) The probability of exactly one change in a sufficiently short interval of length h is approximately λh.

(iii) The probability of two or more changes in a sufficiently short interval is essentially zero. ◘

Remark In this definition, we have modified the usual requirements of a Poisson process by using the words ***approximate Poisson process*** and ***essentially*** in (ii) and (iii) in order to avoid some advanced mathematics. Hence, we refer to this as the ***approximate*** Poisson process.

Suppose that an experiment satisfies the three points of an approximate Poisson process. Let X denote the number of changes in an interval of "length 1" (where "length 1" represents **one unit** of the quantity under consideration). We would like to find an approximation for $P(X = x)$, where x is a **nonnegative integer**. To achieve this, we partition the unit interval into n subintervals of equal length $1/n$. If n is sufficiently large, we shall approximate the probability that x changes occur in this **unit interval** by finding the probability that one change occurs in each of exactly x of these n subintervals. The probability of one occurring in any one subinterval of length $1/n$ is approximately $\lambda(1/n)$ by condition (ii),

$$P(\text{one change occurs in a subinterval}) = \lambda(1/n).$$

The probability of two or more changes in any one subinterval is essentially zero by condition (iii),

$$P(\text{more than one change occurs in a subinterval}) = 0.$$

Hence,
$$P(\text{no changes occur in a subinterval}) = 1 - \lambda(1/n).$$

Consider the occurrence or nonoccurrence of a change in each subinterval as a Bernoulli trial. By condition (i) we have a sequence of n Bernoulli trials with probability p approximate equal to $\lambda(1/n)$. Thus an approximation for $P(X = x)$ is given by the binomial probability

$$\frac{n!}{x!(n-x)!}\left(\frac{\lambda}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^{n-x}.$$

If n increases without bound, we have that

$$\lim_{n\to\infty}\frac{n!}{x!(n-x)!}\left(\frac{\lambda}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^{n-x} = \lim_{n\to\infty}\frac{n(n-1)\cdots(n-x+1)}{n^x}\frac{\lambda^x}{x!}\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-x}.$$

Now, for fixed x, we have

$$\lim_{n\to\infty}\frac{n(n-1)\cdots(n-x+1)}{n^x} = \lim_{n\to\infty}\left[(1)\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{x-1}{n}\right)\right] = 1,$$

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda}, \qquad \text{and} \qquad \lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-x} = 1.$$

Thus
$$\lim_{n\to\infty}\frac{n!}{x!(n-x)!}\left(\frac{\lambda}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^{n-x} = \frac{\lambda^x e^{-\lambda}}{x!} = P(X=x).$$

The distribution of probability associated with this process has a special name. We say that the random variable X has a **Poisson distribution** with the **parameter** $\lambda > 0$, denoted by $X \sim \text{POI}(\lambda)$, if its pmf is of form

$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!}, \qquad x = 0, 1, 2, \ldots.$$

It is easy to see that $f(x)$ enjoys the properties of a pmf since clearly $f(x) \geq 0$ and, from the Maclaurin's series expansion of e^{λ}, we have

$$\sum_{x=0}^{\infty} f(x) = \sum_{x=0}^{\infty}\frac{\lambda^x e^{-\lambda}}{x!} = e^{-\lambda}\sum_{x=0}^{\infty}\frac{\lambda^x}{x!} = e^{-\lambda}e^{\lambda} = 1.$$

The Poisson probability distribution was introduced by S. D. Poisson in 1837. This random variable has a tremendous range of applications in diverse areas since it may be used as an approximation for a binomial random variable with parameters (n, p) when n is large and p is small enough so that np is of a moderate size. In other words, if n independent trials, each of which results in a "success" with probability p, are performed, then, when n is large and p small enough to make np moderate, the number of successes occurring is approximately a Poisson random variable with parameter $\lambda = np$.

Some examples of random variables that usually obey the Poisson probability law follow:

1. The number of misprints on a page (or a group of pages) of a book.
2. The number of people in a community living to 100 years of age.
3. The number of wrong telephone numbers that a dialed in a day.

4. The number of package of dog biscuits sold in a particular store each day.
5. The number of earthquakes occurring during some fixed time span.
6. The number of wars per year.
7. The number of death in a given period of time of the policyholders of a life insurance company.

Example 4.4-1: Suppose that the probability that an item produced by a certain machine will be defective is 0.1. Find the probability that a sample of 10 items will contain at most 1 defective item.

Solution: The desired probability is $\binom{10}{0}(.1)^0(.9)^{10} + \binom{10}{1}(.1)^1(.9)^9 = 0.7361$, whereas the Poisson approximation yields the value $e^{-1} + e^{-1} \approx 0.7358$.

Theorem 4.4-1: If X is a random variable with a Poisson distribution with parameter λ, then

$$\mu = \lambda, \qquad \sigma^2 = \lambda, \qquad \text{and,} \qquad \exp\{\lambda(e^t - 1)\}.$$

Proof: $$E(X) = \sum_{x=0}^{\infty} x \frac{e^{-\lambda}\lambda^x}{x!} = \lambda \sum_{x=1}^{\infty} \frac{e^{-\lambda}\lambda^{x-1}}{(x-1)!} \underbrace{=}_{\text{Let } y=x-1} \lambda \sum_{y=0}^{\infty} \frac{e^{-\lambda}\lambda^y}{y!} = \lambda.$$

$$E[X(X-1)] = \sum_{x=0}^{\infty} x(x-1) \frac{e^{-\lambda}\lambda^x}{x!} = \lambda^2 \sum_{x=2}^{\infty} \frac{e^{-\lambda}\lambda^{x-2}}{(x-2)!} \underbrace{=}_{\text{Let } y=x-2} \lambda^2 \sum_{y=0}^{\infty} \frac{e^{-\lambda}\lambda^y}{y!} = \lambda^2.$$

$$\sigma^2 = E[X(X-1)] + E(X) - [E(X)]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

$$M(t) = E(e^{tX}) = \sum_{x=0}^{\infty} e^{tx} \frac{e^{-\lambda}\lambda^x}{x!} = \sum_{x=0}^{\infty} \frac{e^{-\lambda}(\lambda e^t)^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!} = e^{-\lambda} e^{\lambda e^t}$$

$$= \exp\{\lambda(e^t - 1)\}. \quad \blacksquare$$

4.5 Hypergeometric Distribution

Suppose a population or collection consists of a finite number of items, say *N*, and there are *M* items of type 1 and remaining $N - M$ items are of type 2. Suppose that *n* items are drawn at random ***without replacement***, and denote by *X* the number of items of type 1 that are drawn. The discrete pmf of *X* , called the hypergeometric distribution with parameters *n*, *M*, and *N* and denoted by $X \sim \text{HYP}(n, M, N)$, is given by

$$f(x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}},$$

where *x* is a nonnegative integer subject to the restrictions $x \le n$, $x \le M$, and $n - x \le N - M$. The underlying sample space is taken to be the collection of all subsets of size *n*, of which there are $\binom{N}{n}$, and there are $\binom{M}{x}\binom{N-M}{n-x}$ outcomes that correspond to the event $[X = x]$. According to the **Proposition 1-3**, we have

$$\sum_{x=0}^{n} f(x) = \sum_{x=0}^{n} \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} = 1.$$

Example 4.5-1: A purchaser of electrical components buys them in lots of size 10. It is its policy to inspect 3 components randomly from a lot and to accept the lot only if all 3 are non-defective. If 30 percent of the lots have 4 defective components and 70 percent have only 1, what proportion of lots does the purchaser reject?

Solution: Let A denote the event that the purchaser accepts a lot. Now,

$$P(A) = P(A \mid \text{lot has 4 defectives})(3/10) + P(A \mid \text{lot has 1 defective})(7/10)$$

$$= \frac{\binom{4}{0}\binom{6}{3}}{\binom{10}{3}} \times \frac{3}{10} + \frac{\binom{1}{0}\binom{9}{3}}{\binom{10}{3}} \times \frac{7}{10} = \frac{54}{100}. \quad \blacksquare$$

Theorem 4.5-1: If $X \sim \text{HYP}(n, M, N)$, then for each value *x* = 0, 1,…,n, and as $N \to \infty$ and $M \to \infty$ with $M/N \to p$, a positive constant,

$$\lim_{N\to\infty} \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} = \binom{n}{x} p^x (1-p)^{n-x}.$$

Proof:

$$\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} = \frac{\binom{Np}{x}\binom{N-Np}{n-x}}{\binom{N}{n}} = \frac{(Np)!}{(Np-x)!\,x!} \times \frac{(N-Np)!}{(N-Np-n+x)!\,(n-x)!} \times \frac{n!\,(N-n)!}{N!}$$

Now, $\dfrac{(Np)!}{(Np-x)!} = Np \times (Np-1) \times \cdots \times (Np-(x-1)) = p^x N^x + \text{negligible factors}$

$$\frac{(N-Np)!}{(N-Np-n+x)!} = (N-Np)(N-Np-1)\cdots(N-Np-(n-x-1))$$
$$= N^{n-x}(1-p)^{n-x} + \text{negligible factors}$$

$$\frac{(N-n)!}{N!} = \frac{1}{N(N-1)\cdots(N-(n-1))} = \frac{1}{N^n + \text{negligible factors}}$$

Hence, $\displaystyle\lim_{N\to\infty} \frac{\binom{Np}{x}\binom{N-Np}{n-x}}{\binom{N}{n}} = \frac{n!}{x!\,(n-x)!} p^x (1-p)^{n-x}$. ◻

Example 4.5-2: A lot of 1,000 parts is shipped to a company. A sampling plan dictates that $n = 100$ parts are to be taken at random and ***without replacement*** and the lot accepted if no more two of these 100 parts are defective. Here AC = 2 is usually called the acceptance number. The operating characteristic curve

$$OC(p) = P(X \le 2)$$

where p is the fraction defective in the lot, is really the sum of the three hypergeometric probabilities

$$\sum_{x=0}^{2} f(x) = \sum_{x=0}^{2} \frac{\binom{M}{x}\binom{1000-M}{100-x}}{\binom{1000}{100}}$$

where $M = 1000 \times p$. However we have seen that the hypergeometric distribution can be approximated by the binomial distribution, which in turn can be approximated by the Poisson distribution when n is large and p is small. This exactly our situation since $n = 100$ and we are interested in value of p in the range 0.00 to 0.10. Thus

$$OC(p) = P(X \le 2) \approx \sum_{x=0}^{2} \frac{(100p)^x e^{-100p}}{x!}.$$ ◻

Theorem 4.5-2: If $X \sim \text{HYP}(n, M, N)$, then

$$\mu = n\left(\frac{M}{N}\right) \quad \text{and} \quad \sigma^2 = n\left(\frac{M}{N}\right)\left(\frac{N-M}{N}\right)\left(\frac{N-n}{N-1}\right).$$

Proof: $$E(X) = \sum_{x=0}^{n} x f(x) = \sum_{x=1}^{n} x \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} = \sum_{x=1}^{n} x \frac{\frac{M!}{x!(M-x)!} \times \frac{(N-M)!}{(n-x)!(N-M+x)!}}{\binom{N}{n}}$$

$$= M\sum_{x=1}^{n} \frac{\frac{(M-1)!}{(x-1)!(M-x)!} \times \frac{(N-M)!}{(n-x)!(N-M+x)!}}{\binom{N}{n}}.$$

We now make the change of variables $k = x - 1$ in the summation, and replace

$$\binom{N}{n} \quad \text{with} \quad \frac{N}{n}\binom{N-1}{n-1}$$

in the denominator, this become

$$\mu = \frac{M}{\left(\frac{N}{n}\right)} \sum_{k=0}^{n-1} \frac{\frac{(M-1)!}{k!(M-1-k)!} \times \frac{(N-M)!}{(n-k-1)!(N-M+k+1)!}}{\binom{N-1}{n-1}}$$

$$= n\left(\frac{M}{N}\right)\sum_{k=0}^{n-1} \frac{\binom{M-1}{k}\binom{N-M}{n-1-k}}{\binom{N-1}{n-1}} = n\left(\frac{M}{N}\right).$$

$$E[X(X-1)] = \sum_{x=0}^{n} x(x-1) f(x) = \sum_{x=2}^{n} x(x-1) \frac{\frac{M!}{x!(M-x)!} \times \frac{(N-M)!}{(n-x)!(N-M+x)!}}{\binom{N}{n}}$$

$$= M(M-1)\sum_{x=2}^{n} \frac{\frac{(M-2)!}{(x-2)!(M-x)!} \times \frac{(N-M)!}{(n-x)!(N-M+x)!}}{\binom{N}{n}}.$$

In the summation, let $k = x - 2$, and in the denominator, note that

$$\binom{N}{n} = \frac{N!}{n!(N-n)!} = \frac{N(N-1)}{n(n-1)}\binom{N-2}{n-2}.$$

Thus $$E[X(X-1)] = \frac{M(M-1)}{\frac{N(N-1)}{n(n-1)}} \sum_{k=0}^{n-2} \frac{\binom{M-2}{k}\binom{N-M}{n-2-k}}{\binom{N-2}{n-2}} = \frac{M(M-1)(n)(n-1)}{N(N-1)}.$$

$$\sigma^2 = \frac{M(M-1)(n)(n-1)}{N(N-1)} + \frac{nM}{N} - \left(\frac{nM}{N}\right)^2 = n\left(\frac{M}{N}\right)\left(\frac{N-M}{N}\right)\left(\frac{N-n}{N-1}\right). \quad \square$$

4.6 Probability Generating Functions

An important class of discrete random variables is one in which X represents a count and consequently takes integer values: $X = 0, 1, 2, \ldots$. A Mathematical device useful in finding the probability distributions and other properties of integer-valued random variables is the probability generating function.

Definition 4.6-1: Let X be an integer-valued random variable for which $P(X = i) = p_i$, where $i = 0, 1, 2, \ldots$. The ***probability generating function*** $P(t)$ for X is defined to be

$$P(t) = E\left(t^X\right) = p_0 + p_1 t + p_2 t^2 + \cdots = \sum_{i=0}^{\infty} p_i t^i$$

For all values of t such that $P(t)$ is finite. $\square$

The reason for calling $P(t)$ a probability generating function is that the coefficient of t^i in $P(t)$ is the probability p_i. Repeated differentiation of $P(t)$ yields factorial moments for the random variable X.

Definition 4.6-2: The kth ***factorial moment*** for a random variable X is defined to be

$$\mu_{[k]} = E\left[X(X-1)(X-2)\cdots(X-k+1)\right]$$

where k is positive integer. $\square$

Note that $\mu_{[1]} = E(X) = \mu$. The second factorial moment, $\mu_{[2]} = E[X(X-1)]$, was useful in finding the variance for binomial, geometric, and Poisson random variables.

Theorem 4.6-1: If $P(t)$ is the probability generating function for an integer-valued random variable, X, then the kth factorial moment of X is given by

$$\left.\frac{d^k P(t)}{dt^k}\right]_{t=1} = P^{(k)}(1) = \mu_{[k]}. \quad \square$$

Proof: Since

$$P(t) = p_0 + p_1 t + p_2 t^2 + P_3 t^3 + \cdots$$

it follows that

$$P^{(1)}(t) = \frac{dP(t)}{dt} = p_1 + 2p_2 t + 3p_3 t^2 + 4P_4 t^3 + \cdots,$$

$$P^{(2)}(t) = \frac{d^2P(t)}{dt^2} = (2)(1)p_2 + (3)(2)p_3 t + (4)(3)P_4 t^2 + \cdots,$$

and in general,

$$P^{(k)}(t) = \frac{d^kP(t)}{dt^k} = \sum_{x=k}^{\infty} x(x-1)(x-2)\cdots(x-k+1)p_x t^{x-k} .$$

Setting $t = 1$ in each of these derivatives, we obtain

$$P^{(1)}(1) = p_1 + 2p_2 + 3p_3 + 4P_4 + \cdots = E(X) = \mu_{[1]},$$

$$P^{(2)}(1) = (2)(1)p_2 + (3)(2)p_3 + (4)(3)P_4 + \cdots = E[X(X-1)] = \mu_{[2]},$$

and in general,

$$P^{(k)}(1) = \frac{d^kP(t)}{dt^k} = \sum_{x=k}^{\infty} x(x-1)(x-2)\cdots(y-k+1)p_x = E[X(X-1)\cdots(X-k+1)] = \mu_{[k]} \quad .$$

◘

Example 4.6-1: Find the probability generating function for a geometric random variable. Then use $P(t)$ to find the mean of a geometric random variable.

Solution: Note that $p_0 = 0$ since X cannot assume this value. Then

$$P(t) = E\left(t^X\right) = \sum_{x=1}^{\infty} t^x q^{x-1} p = \sum_{x=1}^{\infty} \frac{p}{q}(qt)^x = \frac{p}{q}\left(\frac{qt}{1-qt}\right) = \frac{pt}{1-qt}, \qquad \text{if } t < \frac{1}{q}.$$

$$E(X) = \mu_{[1]} = P^{(1)}(1) = \frac{d}{dt}\left(\frac{pt}{1-qt}\right)\Bigg]_{t=1} = \frac{(1-qt)p - (pt)(-q)}{(1-qt)^2}\Bigg]_{t=1} = \frac{p^2 + pq}{p^2} = \frac{1}{p}.$$ ◘

Since we already have the moment generating function to assist in finding the moments of a random variable, of what value is $P(t)$? The answer is that it may be difficult to find $M(t)$ but much easier to find $P(t)$. Thus $P(t)$ provides an additional tool for finding the moments of a random variable. It may or may not useful in a given situation.

Finding the moments of a random variable is not the major use of the probability generating function. Its primary application is in deriving the probability function for other integer-valued random variables.