MODULE 4

SOME SPECIAL DISCRETE DISTRIBUTIONS AND THEIR PROPERTIES

LECTURE 19

Topics

4.3 THE HYPERGEOMETRIC DISTRIBUTION

4.4 THE POISSON DISTRIBUTION

4.5 DISCRETE UNIFORM DISTRIBUTION

Example 2.1

Urn U_i $(i = 1, 2)$ contains N_i balls out of which r_i are red and $N_i - r_i$ are black. A sample of n $(1 \leq n \leq N_1)$ balls is chosen at random (without replacement) from urn U_1 and all the balls in the selected sample are transferred to urn U_2. After the transfer two balls are drawn at random from the urn U_2. Find the probability that both the balls drawn from urn U_2 are red.

Solution. Let X denote the number of red balls among the n balls drawn from urn the U_1. Then $X \sim \text{Hyp}(r_1, n, N_1)$. Let E be the even that both the balls drawn from the urn U_2 are red. Then the required probability is

$$P(E) = \sum_{x=0}^{n} P(E|\{X = x\})P(\{X = x\})$$

$$= \sum_{x=\max(0, n-N_1+r_1)}^{\min(n, r_1)} P(E|\{X = x\})P(\{X = x\}).$$

Note that, for $x \in S_X = \{m \in \mathbb{N}: \max(0, n - N_1 + r_1) \leq m \leq \min(n, r_1)\}$,

$$P(E|\{X = x\}) = \frac{\binom{r_2+x}{2}}{\binom{N_2+n}{2}}$$

$$= \frac{r_2(r_2-1)+\ 2r_2x+x(x-1)}{(N_2+n)(N_2+n-1)}.$$

Therefore,

$$P(E) = \frac{1}{(N_2+n)(N_2+n-1)} \sum_{x=\max(0,n-N_1+r_1)}^{\min(n,r_1)} \{r_2(r_2-1)+\ 2r_2x+x(x-1)\}P(\{X=x\})$$

$$= \frac{1}{(N_2+n)(N_2+n-1)}[r_2(r_2-1)+\ 2r_2E(X)+\ E(X(X-1))]$$

$$= \frac{1}{(N_2+n)(N_2+n-1)}\left[r_2(r_2-1)+\ 2r_2\frac{r_1}{N_1}+\ n(n-1)\frac{r_1(r_1-1)}{N_1(N_1-1)}\right].\blacksquare$$

4.4 THE POISSON DISTRIBUTION

Suppose that some event E, say a phone call received at a telephone exchange, is occurring randomly over a period of time and one is interested in the probability distribution of X, the number of times the event E has occurred in an unit interval (say $(0, 1]$). One way to model the probability distribution of X is to partition the unit interval $(0, 1]$ into a large number (say n, where $n \to \infty$) of infinitesimal subintervals $\left(0,\frac{1}{n}\right], \left(\frac{1}{n},\frac{2}{n}\right], \cdots, \left(\frac{n-1}{n}, 1\right]$ of length $\frac{1}{n}$ each. In many situations it may be relevant to assume that:

(i) for each infinitesimal subinterval $\left(\frac{k-1}{n},\frac{k}{n}\right], k = 1, 2, \ldots, n$, the probability that the event E will occur in this subinterval is $\frac{\lambda}{n}$ and the probability that the event E will not occur in this subinterval is $1-\frac{\lambda}{n}$, where $\lambda > 0$ is a given constant;

(ii) chance of two or more occurrences of the event E in any infinitesimal subinterval $\left(\frac{k-1}{n},\frac{k}{n}\right], k = 1, 2, \ldots, n$, is so small that it can be neglected;

(iii) if $\left(\frac{j-1}{n},\frac{j}{n}\right]$ and $\left(\frac{k-1}{n},\frac{k}{n}\right]$ $(1 \le j < k \le n)$ are disjoint subintervals then the number of times the event E occurs in the interval $\left(\frac{j-1}{n},\frac{j}{n}\right]$ is independent of the number of times the event E occurs in the interval $\left(\frac{k-1}{n},\frac{k}{n}\right]$.

Under the above hypotheses, in each infinitesimal subinterval $\left(\frac{k-1}{n}, \frac{k}{n}\right], k = 1, 2, \dots, n,$ event E can occur only 1 or 0 times and the probability of occurrence of event E in each of these subintervals is the same $(= \frac{\lambda}{n})$. If we label the occurrence of event E in any of these subintervals as success (S) and its non-occurrence as failure (F), then we have a sequence of n independent Bernoulli trials with probability of success in each trial as $p_n = \frac{\lambda}{n}$. Therefore it is reasonable to assume that

$$X \equiv X_n = \text{number of times event } E \text{ occurs in the unit interval } (0, 1]$$

$$\sim \mathrm{Bin}(n, p_n),$$

where $p_n = \frac{\lambda}{n}$. The p.m.f. of X ($\equiv X_n$) is given by

$$f_n(k) = \binom{n}{k} p_n^k (1 - p_n)^{n-k} I_{\{0,1,\dots,n\}}(k)$$

$$= \frac{1}{k!}\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right)(np_n)^k \left(1 - \frac{np_n}{n}\right)^n (1 - p_n)^{-k} I_{\{0,1,\dots,n\}}(k).$$

Note that $np_n = \lambda$ and $p_n \to 0$, as $n \to \infty$. Therefore

$$f_n(k) \xrightarrow{n\to\infty} \frac{e^{-\lambda}\lambda^k}{k!} I_{\{0,1,\dots\}}(k)$$

$$= \begin{cases} \frac{e^{-\lambda}\lambda^k}{k!}, & \text{if } k \in \{0, 1, \dots\} \\ 0, & \text{otherwise} \end{cases}. \tag{3.1}$$

Note that $\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{\lambda}$, and therefore (3.1) defines a probability mass function. The above discussion suggests that the probability distribution of X, the number of times the event E occurs in the unit interval $(0, 1]$, can be modeled by the p.m.f.

$$f_X(k) = P(\{X = k\}) = \begin{cases} \frac{e^{-\lambda}\lambda^k}{k!}, & \text{if } k \in \{0, 1, \dots\} \\ 0, & \text{otherwise} \end{cases},$$

where $\lambda > 0$ may be interpreted as the intensity with which event E occurs. ■

Definition 3.1

A discrete type random variable X is said to follow a *Poisson distribution* with parameter (or intensity) $\lambda > 0$ (written as $X \sim \mathrm{P}(\lambda)$) if its support is $S_X = \{0, 1, 2, \dots\}$ and its probability mass function is given by

$$f_X(k) = P(\{X = k\}) = \begin{cases} \dfrac{e^{-\lambda}\lambda^k}{k!}, & \text{if } k \in \{0, 1, \dots\} \\ 0, & \text{otherwise} \end{cases}. \blacksquare$$

Notice that we have a family $\{P(\lambda) : \lambda > 0\}$ of Poisson distributions corresponding to various choices of $\lambda > 0$. Also notice that a Poisson distribution can be derived as a limiting binomial distribution. In fact on exactly following the arguments used for deriving (3.1) we can prove the following theorem.

Theorem 3.1

Let $X_n \sim \text{Bin}(n, p_n), n = 1, 2, \dots,$ where $\{p_n : n = 1, 2, \dots\}$ is a sequence of real numbers in $(0, 1)$ such that $\lim_{n\to\infty}(np_n) = \lambda$, for some real constant $\lambda > 0$. Let $f_{X_n}(x)$ denote the p.m.f. of X_n. Then

$$\lim_{n\to\infty} f_{X_n}(k) = f(k) = \begin{cases} \dfrac{e^{-\lambda}\lambda^k}{k!}, & \text{if } k \in \{0, 1, \dots\} \\ 0, & \text{otherwise} \end{cases}. \blacksquare$$

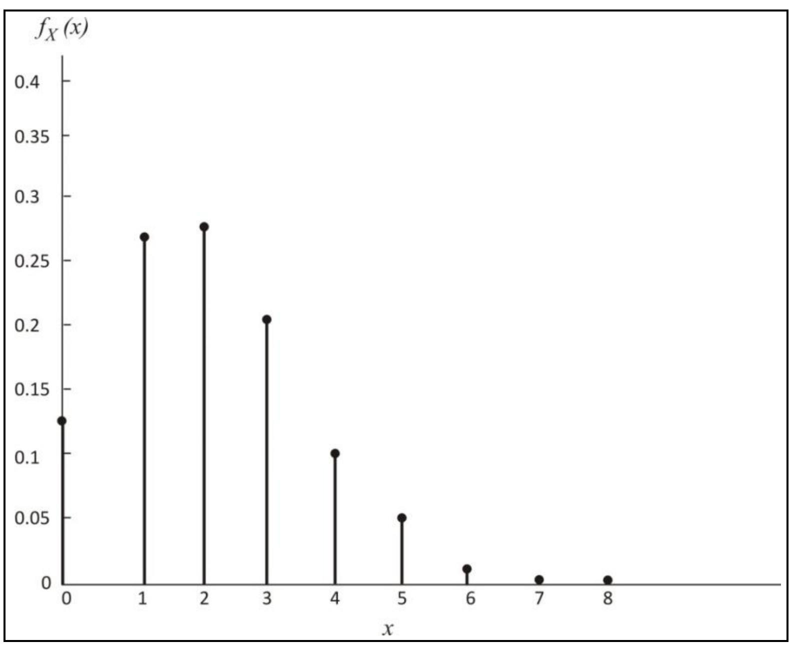

Figure 3.1. Plot of p.m.f. of Bin(50,0.04)

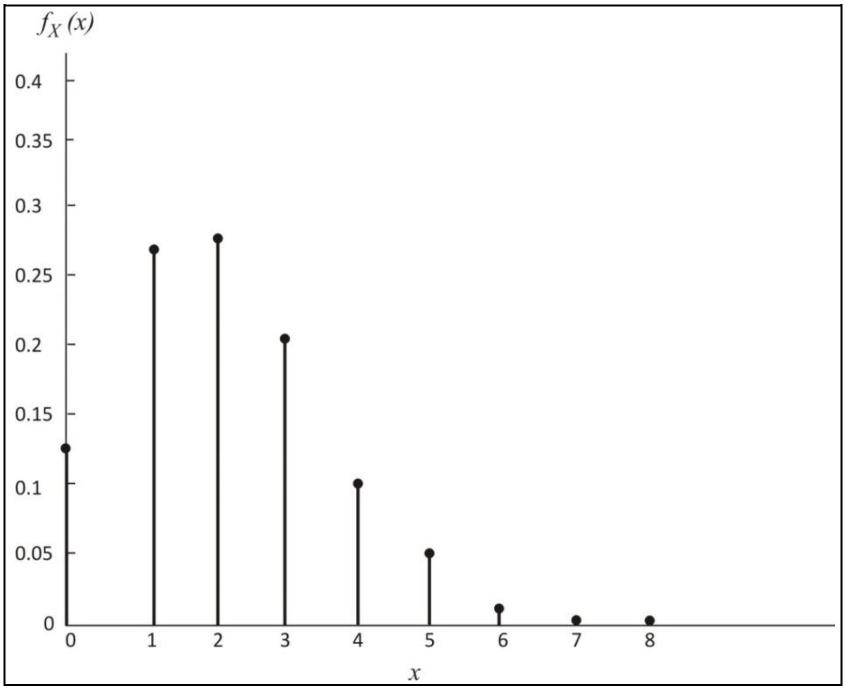

Figure 3.2. Plot of p.m.f. of P(2)

The above theorem suggests that, in situations where we have a large number (say n) of independent Bernoulli trials and probability of success (say p) in each trial is so small that np is a moderate quantity, the Poisson distribution provides a good fit (see Figures 3.1 and 3.2).

Example 3.1

Consider a person who plays a series of 2500 games independently. If the probability of person winning any game is 0.002, find the probability that the person will win at least two games.

Solution. In each game let us label winning of the game by the person as a success and his/her losing the game as a failure. Then we have a sequence of $n = 2500$ Bernoulli trials with probability of success in each trial as $p = 0.002$. Let X denote the number of successes in $n = 2500$ Bernoulli trials. Then $X \sim \text{Bin}(2500, 0.002)$ and the desired probability is

$$P(\{X \geq 2\}) = 1 - \{P(\{X = 0\}) + P(\{X = 1\})\}$$

$$= 1 - \{(1 - 0.002)^{2500} + 2500 \times 0.002 \times (1 - 0.002)^{2499}\}.$$

Since $n = 2500$ is large and $np = 5$ is a moderate quantity, we can approximate the distribution of $X \sim \text{Bin}(n, p)$ by that of $Y \sim \text{P}(\lambda)$, where $\lambda = np = 5$. Therefore

$$P(\{X \geq 2\}) \approx P(\{Y \geq 2\})$$
$$= 1 - \{P(\{Y = 0\}) + P(\{Y = 1\})\}$$
$$= 1 - \left(e^{-\lambda} + \lambda e^{-\lambda}\right)$$
$$= 1 - 6 \times e^{-5}$$
$$= 1 - 0.0404$$
$$= 0.9596,$$

where the symbol ≈ stands for "approximately equal to". ■

Example 3.2

Telephone calls arrive independently at a telephone exchange according to Poisson distribution with mean rate of 5 calls per second. Find the probability that:

(i) three calls will be received during a second;
(ii) among first 5 one second time intervals $(0,1], (1,2], \ldots, (4,5]$ of a given minute, exactly 3 calls will be received on 2 of these time intervals.

Solution.

(i) Let X denote the number of calls received during a second. Then $X \sim P(5)$ and the desired probability is

$$P(\{X = 3\}) = \frac{e^{-5}5^3}{3!} \approx 0.1404.$$

(ii) On any given interval $(k-1,k], k = 1,2,\ldots,5$, let us label arrival of 3 calls as success and arrival of any other number of calls as failure. Then we have a sequence of 5 independent Bernoulli trials with probability of success in each trial as 0.1404 (see (i)). Let Y denote the number of one-second time intervals among $(0,1], (1,2], \ldots, (4,5]$ on which exactly three calls are received (i.e., Y denotes the number of successes in 5 independent Bernoulli trials with probability of success in each trial as 0.1404). Then $Y \sim \mathrm{Bin}(5, 0.1404)$ and the desired probability is

$$P(\{Y = 2\}) = \binom{5}{2}(0.1404)^2(1 - 0.1404)^3 \approx 0.1252. \ ■$$

Suppose that $X \sim \mathrm{P}(\lambda)$, for some $\lambda > 0$. Then, for $r \in \{1,2,\ldots\}$, the r-th factorial moment of X is given by

$$E\left(X_{(r)}\right) = E\left(X(X-1)\cdots(X-r+1)\right)$$

$$= e^{-\lambda} \sum_{k=0}^{\infty} k(k-1)\cdots(k-r+1)\frac{\lambda^k}{k!}$$

$$= e^{-\lambda} \sum_{k=r}^{\infty} \frac{\lambda^k}{(k-r)!}$$

$$= e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^{k+r}}{k!}$$

$$= \lambda^r e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}$$

$$= \lambda^r.$$

Therefore

$$\boxed{\text{Mean } = E(X) = E\left(X_{(1)}\right) = \lambda;}$$

$$E(X(X-1)) = \lambda^2;$$

and
$$\boxed{\text{Var}(X) = E(X(X-1)) + E(X) - \left(E(X)\right)^2 = \lambda.}$$

Thus, for the Poisson distribution, the mean is equal to the variance

The m.g.f. of $X \sim P(\lambda)$ is given by

$$M_X(t) = E(e^{tX})$$

$$= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^t)^k}{k!}$$

$$= e^{-\lambda} \mathrm{e}^{\lambda e^t}$$

$$= e^{\lambda(e^t - 1)}, \qquad t \in \mathbb{R}.$$

Therefore

$$\Psi_X(t) = \ln M_X(t) = \lambda(e^t - 1), \quad t \in \mathbb{R}.$$

and
$$\Psi_X^{(1)}(t) = \Psi_X^{(2)}(t) = \lambda e^t, \quad t \in \mathbb{R}.$$

Consequently,

$$E(X) = \Psi_X^{(1)}(0) = \lambda$$

and $$\mathrm{Var}(X) = \Psi_X^{(2)}(0) = \lambda.$$

The Poisson distribution provides a good model for counts of an event over a period of time, over an area or over a volume.

4.5 DISCRETE UNIFORM DISTRIBUTION

For a given positive integer $N\ (\geq 2)$ and real numbers $x_1 < x_2 < \cdots < x_N$, a r.v. X of discrete type is said to follow a *discrete uniform distribution* on the set $\{x_1, x_2, \dots, x_N\}$ (written as $X \sim U(\{x_1, x_2, \dots, x_N\})$ if the support of X is $S_X = \{x_1, x_2, \dots, x_N\}$ and its p.m.f. is given by

$$f_X(x) = P(\{X = x\}) = \begin{cases} \frac{1}{N}, & \text{if } x \in S_X = \{x_1, x_2, \dots, x_N\} \\ 0, & \text{otherwise} \end{cases}.$$

Suppose that $X \sim U(\{x_1, x_2, \dots, x_N\})$. Then

$$\mu_r' = E(X^r) = \frac{1}{N}\sum_{i=1}^{N} x_i^r, \qquad r \in \{1, 2, \dots\},$$

$$\text{Mean} = \mu_1' = E(X) = \frac{1}{N}\sum_{i=1}^{N} x_i$$

and $$\mathrm{Var}(X) = \sigma^2 = E\left(\left(X - \mu_1'\right)^2\right) = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \mu_1'\right)^2.$$

Also the m.g.f. of $X \sim U(\{x_1, x_2, \dots, x_N\})$ is given by

$$M_X(t) = E(e^{tX}) = \frac{1}{N}\sum_{i=1}^{N} e^{tx_i}, \quad t \in \mathbb{R}.$$

Now suppose that $Y \sim U(\{1, 2, \dots, N\})$. Then

$$\boxed{E(Y) = \frac{1}{N}\sum_{i=1}^{N} i = \frac{N+1}{2},}$$

$$E(Y^2) = \frac{1}{N}\sum_{i=1}^{N} i^2 = \frac{(N+1)(2N+1)}{6}$$

and

$$\boxed{\mathrm{Var}(Y) = E(Y^2) - \left(E(Y)\right)^2 = \frac{N^2-1}{12}.}$$

Also the m.g.f. of $Y \sim U(\{1,2,\dots,N\}$ is given by

$$M_Y(t) = E(e^{tY}) = \frac{1}{N}\sum_{i=1}^{N} e^{it} = \begin{cases} \dfrac{e^t(e^{Nt}-1)}{e^t-1}, & \text{if } t \neq 0 \\ 1, & \text{if } t = 0 \end{cases}. \blacksquare$$

Example 4.1

A person has to open a lock whose key is lost among a set of N keys. Assume that out of these N keys only one can open the lock. To open the lock the person tries keys one by one by choosing, at each attempt, one of the keys at random from the unattempted keys. The unsuccessful keys are not considered for future attempts. Let Y denote the number of attempts the person will have to make to open the lock. Show that $Y \sim U(\{1,2,\dots,N\})$ and hence find the mean and the variance of the r.v. Y.

Solution. For $r \notin \{1,2,\dots,N\}$, we have $P(\{Y=r\}) = 0$. For $r \in \{1,2,\dots,n\}$

$$P(\{Y=r\}) = \frac{N-1}{N}\cdot\frac{N-2}{N-1}\cdots\frac{N-(r-1)}{N-(r-2)}\cdot\frac{1}{N-(r-1)} = \frac{1}{N}.$$

It follows that $Y \sim U(\{1,2,\dots,N\})$ and, therefore,

$$\boxed{E(Y) = \frac{N+1}{2} \text{ and } \mathrm{Var}(Y) = \frac{N^2-1}{12}. \blacksquare}$$