IC105: Probability and Statistics	**2021-22-M**

Lecture 12: Some Special Discrete Distributions

Instructor: Dr. Kuldeep Kumar Kataria *Scribe:*

12.1. Some Special Discrete Distributions

12.1.1. Bernoulli and Binomial Distribution

Bernoulli Experiment: A random experiment with just two possible outcomes (say success (S) and failure (F)). Each replication of a Bernoulli experiment is called a Bernoulli trial.

Consider a sequence of n independent Bernoulli trials with probability of success (S) in each trial as $p \in (0,1)$ (same for each trial); here $n \in \mathbb{N}$ is a fixed natural number.

Define $X =$ the number of success in n trials. Then $S_X = \{0, 1, 2, \ldots, n\}$ and for $k \in S_X$

$$\begin{aligned}
P(X = k) &= P(\underbrace{SS\cdots SFF\cdots F}_{k \text{ successes and } n-k \text{ failures}}) + P(\underbrace{SFFS\cdots FFS}_{k \text{ successes and } n-k \text{ failures}}) + \cdots + P(\underbrace{FF\cdots FSS\cdots S}_{k \text{ successes and } n-k \text{ failures}}) \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \left(\text{total of } \binom{n}{k} \text{ terms}\right) \\
&= p^k(1-p)^{n-k} + p^k(1-p)^{n-k} + \cdots + p^k(1-p)^{n-k} \\
&= \binom{n}{k} p^k (1-p)^{n-k}, \quad (\text{independence of trials}).
\end{aligned}$$

Thus,

$$f_X(x) = P(X = x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & x = 0, 1, 2, \ldots, n, \\ 0, & \text{otherwise}, \end{cases}$$

is called Binomial distribution with n trials and success probability p (denoted by $Bin(n, p)$ and written as $X \sim Bin(n, p)$). $\{Bin(n, p) : n \in \mathbb{N}, p \in (0, 1)\}$ is the family of probability distributions that has two parameters $n \in \mathbb{N}$ and $p \in (0, 1)$.

$\{Bin(1, p) : p \in (0, 1)\}$: Bernoulli distributions. $Bin(1, p)$: Bernoulli distribution with success probability $p \in (0, 1)$.

Suppose that $X \sim Bin(n, p)$, $n \in \mathbb{N}, p \in (0, 1)$. Then

$$\text{m.g.f. } M_X(t) = E(e^{tX}) = \sum_{k=0}^{\infty} e^{tk} \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=0}^{\infty} \binom{n}{k} (pe^t)^k (1-p)^{n-k} = (1 - p + pe^t)^n, \quad t \in \mathbb{R}.$$

Let $q = 1 - p$, so that $M_X(t) = (q + pe^t)^n$, $t \in \mathbb{R}$. Then

$$\begin{aligned}
M_X^{(1)}(t) &= n(q + pe^t)^{n-1} pe^t, \\
M_X^{(2)}(t) &= np(q + pe^t)^{n-1} e^t + n(n-1)(q + pe^t)^{n-2}(pe^t)^2, \\
E(X) &= M_X^{(1)}(0) = np, \;\; E(X^2) = M_X^{(2)}(0) = np + n(n-1)p^2, \;\; \mathrm{Var}(X) = np(1-p) = npq.
\end{aligned}$$

Note that if $X \sim Bin(n,p)$ then Variance $<$ Mean. It can be seen that

$$\begin{aligned}
\mu_3' &= E(X^3) = np(1 - 3p + 3np + 2p^2 - 3np^2 + n^2p^2), \\
\mu_4' &= E(X^4) = np(1 - 7p + 7np + 12p^2 - 18np^2 + 6n^2p^2 - 6p^3 + 11np^3 - 6n^2p^3 + n^3p^3), \\
\mu_3 &= E((X - \mu_1')^3) = np(1-p)(1-2p), \\
\mu_4 &= E((X - \mu_1')^4) = np(1-p)(3p^2(2-n) + 3p(n-2) + 1), \\
\beta_1 &= \frac{\mu_3}{\mu_2^{3/2}} = \frac{1-2p}{\sqrt{np(1-p)}} = \begin{cases} \text{symmetric for } p = \frac{1}{2} \;\;, \\ \text{positively skewed for } 0 < p < \frac{1}{2}, \\ \text{negatively skewed for } p > \frac{1}{2}, \end{cases} \\
\nu_2 &= \nu_1 - 3 = \frac{1 - 6pq}{npq}, \;\; \text{where } \nu_1 = \frac{\mu_4}{\mu_2^2}.
\end{aligned}$$

Also, for $r \in \{1, 2, \dots\}$, let $X_{(r)} = X(X-1)(X-2)\cdots(X-r+1)$, the rth factorial moment is given by

$$\begin{aligned}
E(X_{(r)}) &= \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} k(k-1)(k-2)\cdots(k-r+1) \\
&= n(n-1)(n-2)\cdots(n-r+1) \sum_{k=r}^{n} \binom{n-r}{k-r} p^k (1-p)^{n-k} \\
&= n(n-1)(n-2)\cdots(n-r+1) p^r \sum_{k=0}^{n-r} \binom{n-r}{k} p^k (1-p)^{n-r-k} \\
&= n(n-1)(n-2)\cdots(n-r+1) p^r (1 - p + p)^{n-r} = n(n-1)(n-2)\cdots(n-r+1)p^r.
\end{aligned}$$

Theorem 12.1. *Let $X_1, X_2, \dots, X_k$ be independent r.v.'s with $X_i \sim Bin(n_i, p)$, $n_i \in \mathbb{N}$, $p \in (0,1)$, $i = 1, 2, \dots, k$. Then $Y = \sum_{i=1}^k X_i \sim Bin(n, p)$, where $n = \sum_{i=1}^k n_i$.*

Proof. For $t \in \mathbb{R}$,

$$\begin{aligned}
M_Y(t) = E(e^{tY}) = E(e^{t\sum_{i=1}^k X_i}) = E\left(\prod_{i=1}^k e^{tX_i}\right) &= \prod_{i=1}^k E(e^{tX_i}), \quad \text{(independent of } X_i's) \\
&= \prod_{i=1}^k M_{X_i}(t) = \prod_{i=1}^k (1 - p + pe^t)^{n_i} = (1 - p + pe^t)^{\sum_{i=1}^k n_i} \\
&\to \text{m.g.f. of } Bin\left(\sum_{i=1}^k n_i, p\right).
\end{aligned}$$

By uniqueness of m.g.f. $Y \sim Bin(n,p)$, where $n = \sum_{i=1}^k n_i$. $\square$

Example 12.2. *Let $X \sim Bin(n, 1/2)$, then $X - \frac{1}{2} \stackrel{d}{=} \frac{n}{2} - X$, since $n - X \stackrel{d}{=} X$ (Exercise).*

Example 12.3. *A fair dice is rolled* 5 *times independently. Find the probability that on* 3 *occasions we get a six.*

Solution: Consider getting a six as success. Then $X =$ the number of success in 5 trials $\sim Bin(5, 1/6)$.

So, the required probability =$P(X = 3) = \binom{5}{3}(1/6)^3(5/6)^2$.

12.1.2. Negative Binomial Distribution

Consider a sequence of independent Bernoulli trials with probability of success in each trial as $p \in (0, 1)$. Let $r \in \{1, 2, \dots\}$ be a fixed positive integer. Let X denote the number of failures preceding the rth success. Then $S_X = \{0, 1, 2, \dots\}$ and for $k \in S_X$, we have

$$\begin{aligned}
f_X(k) &= P(X = k) \\
&= P(k \text{ failures precede } r\text{th success}) \\
&= P(r-1 \text{ successes in first } k+r-1 \text{ trials and success in } (k+r)\text{th trial}) \\
&= P(r-1 \text{ successes in first } k+r-1 \text{ trials}) \times P(\text{success in } (k+r)\text{th trial}), \quad \text{(independence of trials)} \\
&= \binom{k+r-1}{r-1} p^{r-1}(1-p)^k p = \binom{k+r-1}{r-1} p^r (1-p)^k.
\end{aligned}$$

Thus,

$$f_X(x) = P(X = x) = \begin{cases} \binom{x+r-1}{r-1} p^r (1-p)^x, & x = 0, 1, 2, \dots, \\ 0, & \text{otherwise.} \end{cases}$$

The distribution of X is called Negative binomial distribution with r success, and success probability $p \in (0, 1)$ (denoted by $NB(r, p)$ and written as $X \sim NB(r, p)$) (has two parameters $r \in \mathbb{N}$ and $p \in (0, 1)$). $\{NB(r, p) : r \in \mathbb{N},\ p \in (0, 1)\}$ is a family of probability distribution.

Remark 12.4. *For* $t \in (-1, 1)$*, we have*

$$\sum_{k=0}^{\infty} \binom{k+r-1}{r-1} t^k = 1 + rt + \frac{(r+1)r}{2!} t^2 + \frac{(r+2)(r+1)r}{3!} t^3 + \cdots = (1-t)^{-r}.$$

The m.g.f. of $X \sim NB(r, p)$ is

$$\begin{aligned}
M_X(t) = E(e^{tX}) &= \sum_{k=0}^{\infty} e^{tk} \binom{k+r-1}{r-1} (1-p)^k p^r \\
&= p^r \sum_{k=0}^{\infty} \binom{k+r-1}{r-1} ((1-p)e^t)^k = \left(\frac{p}{1-(1-p)e^t} \right)^r, \quad t < -\ln(1-p).
\end{aligned}$$

Thus,

$$\begin{aligned}
&\psi_X(t) = \ln M_X(t) = r \ln p - r \ln(1 - qe^t),\ t < -\ln(1-p), \\
&\psi_X^{(1)}(t) = \frac{rqe^t}{1-qe^t} = r\left(\frac{1}{1-qe^t} - 1 \right), \quad t < -\ln(1-p), \\
&\psi_X^{(2)}(t) = \frac{rqe^t}{(1-qe^t)^2}, \quad t \in \mathbb{R}, \\
&E(X) = \psi_X^{(1)}(0) = \frac{rq}{p}, \quad \mathrm{Var}(X) = \psi_X^{(2)}(0) = \frac{rq}{p^2}, \quad \text{Variance} > \text{Mean.}
\end{aligned}$$

Also, for $m \in \{1, 2, \dots\}$, let $X_{(m)} = X(X-1)(X-2)\cdots(X-m+1)$. Then

$$\begin{aligned}
E(X_{(m)}) &= \sum_{k=0}^{\infty} k(k-1)(k-2)\cdots(k-m+1)\binom{k+r-1}{r-1}p^r(1-p)^k \\
&= p^r \sum_{k=m}^{\infty} k(k-1)(k-2)\cdots(k-m+1)\binom{k+r-1}{r-1}(1-p)^k \\
&= r(r+1)(r+2)\cdots(r+m-1)p^r \sum_{k=m}^{\infty} \frac{(k+r-1)!}{(k-m)!(r+m-1)!}(1-p)^k \\
&= r(r+1)(r+2)\cdots(r+m-1)p^r \sum_{k=0}^{\infty} \frac{(k+m+r-1)!}{k!(r+m-1)!}(1-p)^{k+m} \\
&= r(r+1)(r+2)\cdots(r+m-1)p^r q^m \sum_{k=0}^{\infty} \binom{k+m+r-1}{m+r-1}q^k \\
&= r(r+1)(r+2)\cdots(r+m-1)p^r q^m (1-q)^{-(m+r)} = r(r+1)(r+2)\cdots(r+m-1)(q/p)^m.
\end{aligned}$$

$$\mu_1' = E(X) = \frac{rq}{p}; \quad \mu_2' = E(X^2) = \frac{rq(1+rq)}{p^2}.$$

It can be seen that

$$\begin{aligned}
\mu_3' &= E(X^3) = \frac{q(rp^2 + 3pqr + q^2 r(r+1))}{p^3}, \\
\mu_4' &= E(X^4) = \frac{q(rp^3 + 7p^2 qr + 6pq^2 r(r+1) + q^3 r(r+1)(r+2))}{p^4}, \\
\mu_2 &= E((X-\mu_1')^2) = r(1-p), \\
\mu_3 &= E((X-\mu_1')^3) = \frac{r(p-1)(p-2)}{p^3}, \\
\mu_4 &= E((X-\mu_1')^4) = \frac{r(1-p)(6-6p+p^2+3r-3pr)}{p^4}, \\
\beta_1 &= \frac{\mu_3^2}{\mu_2^3} = \frac{2-p}{\sqrt{rq}} > 0 \quad \text{(positively skewed)}, \\
\nu_2 &= \nu_1 - 3 = \frac{p^2-2p+6}{rq}, \ \text{ where } \nu_1 = \frac{\mu_4}{\mu_2^2}.
\end{aligned}$$

$NB(1, p)$ distribution is called a geometric distribution (denoted by $Ge(p)$, $0 < p < 1$). The p.m.f. of $Y \sim Ge(p)$ is given by

$$f_Y(y) = P(Y = y) = \begin{cases} pq^y, & y = 0, 1, 2, \dots, \\ 0, & \text{otherwise.} \end{cases}$$

$P(Y \geq m) = p\sum_{y=m}^{\infty} q^y = q^m$. This implies that

$$\begin{aligned}
P(Y \geq m+n | Y \geq m) &= \frac{P(Y \geq m+n, Y \geq m)}{P(Y \geq m)} = \frac{P(Y \geq m+n)}{P(Y \geq m)} \\
&= \frac{q^{m+n}}{q^m} = q^n = P(Y \geq n), \ \forall\, m, n \in \{0, 1, \dots\}. \qquad (12.1)
\end{aligned}$$

Also,

$$P(Y \geq m+n) = P(Y \geq m)P(Y \geq n), \ \forall\, m, n \in \{0, 1, \dots\}. \qquad (12.2)$$

Remark 12.5. *The property (12.1) possessed by $Ge(p)$ distribution has an interesting interpetation. Suppose that a device can absorb $0, 1, 2, \dots$ shocks before failing. Let T denote the random variable representing the number of shocks that device can absorb before failing.*

$P(T \geq m+n | T \geq m)$: *conditional probability that a system has absorbed m shocks will absorb atleast n additional shocks before failing.*

$P(T \geq n)$: *a new device can survive atleast n shocks before failing.*

Thus if distribution of T has property (12.1) then the age of the device has no effect as the residual (remaining) life of the device (implying that an used device is as good as a new device). The property (12.1) (or equivalently (12.2)) is famously known as Lack of memory (LoM) property.

Theorem 12.6. *Let T be a discrete type r.v. with range $S_T = \{0, 1, 2, \dots\}$. Then T has the lack of memory property if and only if $T \sim Ge(p)$, for some $p \in (0, 1)$.*

Proof. Obviously, $T \sim Ge(p)$, for some $p \in (0,1) \implies T$ has LoM property. Then $P(T \geq j+k) = P(T \geq j)P(T \geq k)\ \forall\, j,k \in \{0, 1, \dots\}$. Let $P(T = 0) = p$. Then $p \in (0, 1)$ and for $j \in \{0, 1, \dots\}$

$$\begin{aligned} P(T \geq j+1) &= P(T \geq j)P(T \geq 1) \\ &= P(T \geq j)(1-p) \\ &= P(T \geq j-1)(1-p)^2 \\ &\vdots \\ &= P(T \geq 0)(1-p)^{j+1} = (1-p)^{j+1} \end{aligned}$$

This implies

$$P(T = k) = P(T \geq k) - P(T \geq k+1) = p(1-p)^k,\ \ k = \{0, 1, 2, \dots\} \implies T \sim Ge(p).$$

This completes the proof. □

Example 12.7. *A person repeatedly rolls a fair die independently untill an upper face with two or three dots is observed twice. Find the probability that the person would require eight rolls to achive this.*

Solution: Consider getting 2 or 3 dots as success. Let $Z =$ the number of trials requires to get 2 successes. Then probability of success in each trial is $1/3$ and required probability $= P(Z = 8) = \left\{\binom{7}{1}\frac{1}{3}\left(\frac{2}{3}\right)^6\right\} \times \frac{1}{3} = \frac{448}{6561}$.

12.1.3. The Hypergeometric Distribution

Consider a population comprising of N (≥ 2) units out of which $a \in \{1, 2, \dots, N-1\}$ are labelled as S (success) and $N - a$ are labeled as F (failure). A sample of size n is drawn from this population drawing one unit at a time. Let X denotes the number of successes in drawn sample.

Case-I: Drawn are independent and sampling is with replacement (*i.e.* after each draw the drawn units is replaced back into the population)

In this case we have sequence of n independent Bernoulli trials with probability of success in each trial as $p = \frac{a}{N}$. Thus $X \sim Bin(n, \frac{a}{N})$.

Case-II: Without replacement (*i.e.* drawn units are not replaced back into the population).

Here,

$$
\begin{aligned}
&P(\text{obtaining } S \text{ in first draw}) = \frac{a}{N},\\
&P(\text{obtaining } S \text{ in second draw}) = \frac{a}{N}\frac{a-1}{N-1} + \frac{N-a}{N}\frac{a}{N-1} = \frac{a}{N}.\\
&\text{In general, } P(\text{obtaining } S \text{ in } i\text{th trial}) = \frac{a}{N},\ \ i = 1, 2, \dots, n \text{ (Exercise)},\\
&P(\text{obtaining } S \text{ in first and second trial}) = \frac{a}{N}\frac{a-1}{N}\\
&\qquad\qquad \neq \frac{a}{N}\frac{a}{N} = P(\text{obtaining } S \text{ in first trial}) \times P(\text{obtaining } S \text{ in second trial})\\
&\qquad\qquad \implies \text{Draws are not independent.}
\end{aligned}
$$

Thus, we can not conclude that $X \sim Bin(n, \frac{a}{N})$. So,

$$
f_X(x) = P(X = x) = \begin{cases} \frac{\binom{a}{x}\binom{N-a}{n-x}}{\binom{N}{n}}, & x = \max\{0, n-N+a\}, \dots, \min\{n, a\},\\ 0, & \text{otherwise.} \end{cases}
$$

The distribution is called Hypergeometric distribution (Hyp(a, n, N)). It has three parameters $N \in \{2, 3, \dots\}$, $a, n \in \{1, 2, \dots, N-1\}$.

For $r \in \mathbb{N}$, let $X_{(r)} = X(X-1)(X-2)\cdots(X-r+1)$. Then

$$
E(X_{(r)}) = \frac{1}{\binom{N}{n}} \sum_{k=\max\{0,n-N+a\}}^{\min\{n,a\}} k(k-1)(k-2)\cdots(k-r+1)\binom{a}{k}\binom{N-a}{n-k}.
$$

Clearly for $r > \min\{n, a\}$, $E(X_{(r)}) = 0$. For $1 \le r \le \min\{n, a\}$, we have

$$
\begin{aligned}
E(X_{(r)}) &= \frac{1}{\binom{N}{n}} \sum_{k=\max\{r,n-N+a\}}^{\min\{n,a\}} k(k-1)(k-2)\cdots(k-r+1)\binom{a}{k}\binom{N-a}{n-k}\\
&= \frac{a_{(r)}}{\binom{N}{n}} \sum_{k=\max\{r,n-N+a\}}^{\min\{n,a\}} \binom{a-r}{k-r}\binom{N-a}{n-k}\\
&= \frac{a_{(r)}}{\binom{N}{n}} \sum_{k=\max\{0,n-N+a-r\}}^{\min\{n-r,a-r\}} \binom{a-r}{k}\binom{N-a}{n-r-k}\\
&= \frac{a_{(r)}}{\binom{N}{n}} \sum_{k=\max\{0,(n-r)-(N-r)+a-r\}}^{\min\{n-r,a-r\}} \binom{a-r}{k}\binom{(N-r)-(a-r)}{(n-r)-k} = \frac{\binom{N-r}{n-r}}{\binom{N}{n}} a_{(r)}.
\end{aligned}
$$

Since $\sum_{k=\max\{0,m-M+b\}}^{\min\{m,b\}} \binom{b}{k}\binom{M-b}{m-k} = \binom{M}{m}$. Thus, for $r \in \mathbb{N}$, we have

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

In particular,

$$\begin{aligned}
E(X) &= E(X_{(1)}) = n\frac{a}{N} = np \text{ (say), where } p = \frac{a}{N},\\
E(X(X-1)) &= E(X_{(2)}) = \frac{n(n-1)}{N(N-1)}a(a-1),\\
\mathrm{Var}(X) &= E(X^2) - (E(X))^2\\
&= E(X(X-1)) + E(X) - (E(X))^2\\
&= n\frac{a}{N}(1-\frac{a}{N})\frac{N-n}{N-1} = np(1-p)\left(1-\frac{n-1}{N-1}\right). \qquad (12.3)
\end{aligned}$$

Remark 12.8. *In case of sampling with replacement we have $X \sim Bin(n,p)$, $E(X) = np$ and $\mathrm{Var}(X) = np(1-p)$, where $p = \frac{a}{N}$. The factor $(1-\frac{n-1}{N-1})$ which on multiplying to variance of $Bin(n,p)$ distribution yields the variance of $Hyp(a,n,N)$ distribution (see 12.3) is called the finite population correction (f.p.c.). Clearly if the sample size n is significantly smaller than the population size N ($n << N$) then f.p.c. will be close to 1 and variance of $Bin(n,p)$ and $Hyp(a,n,N)$ distribution will be very close. Infact when $n << N$ and $n << a \equiv a_N$ (say) are such that $\frac{a_N}{N}$ is a fixed quantity (i.e. as $N \to \infty$, $a_N \to \infty$ and $\frac{a_N}{N} \to p \in (0,1)$, where $p \in (0,1)$ is a fixed quantity) then $Bin(n, \frac{a}{N})$ distribution provides an approximation to $Hyp(a,n,N)$ distribution. 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.*

Theorem 12.9 (Binomial Approximation to Hypergeometric Distributon)**.** *Let $X_{a_N,n,N} \sim Hyp(a_N,n,N)$, where a_N depedns 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*

$$\lim_{N\to\infty} f_{a_N,n,N}(k) = \lim_{N\to\infty} P(X_{a_N,n,N} = k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k}, & k \in \{0,1,2,\ldots,n\},\\ 0, & \textit{otherwise,} \end{cases}$$

i.e., for large N and large a_N, so that $p = \frac{a_N}{N} \in (0,1)$ is a fixed quantity, $Hyp(a_N,n,N)$ probabilities can be approximated by $Bin(n, \frac{a}{N})$ probabilities.

Proof. $S_X = \{m \in \mathbb{N} : \max\{0, n-N+a_N\} \le m \le \min\{n, a_N\}\}$, $n - N + a_N = N(\frac{n}{N} - 1 + \frac{a_N}{N}) \to \infty$ and $a_N = N\frac{a_N}{N} \to \infty$, as $N \to \infty$. Also for $k \in S_X$,

$$\begin{aligned}
f_X(k) &= \binom{n}{k}\left\{\prod_{j=0}^{k-1}\left(\frac{a_N-j}{N-j}\right)\right\}\left\{\prod_{j=0}^{n-k-1}\left(\frac{N-a_N-j}{N-j}\right)\right\}\\
&\xrightarrow{N\to\infty} \binom{n}{k}\left\{\prod_{j=0}^{k-1}(p)\right\}\left\{\prod_{j=0}^{n-k-1}(1-p)\right\} = \binom{n}{k}p^k(1-p)^{n-k}\\
\implies \lim_{N\to\infty} f_{a_N,n,N}(k) &= \begin{cases} \binom{n}{k} p^k(1-p)^{n-k}, & k \in \{0,1,2,\ldots,n\}\\ 0, & \text{otherwise.} \end{cases}
\end{aligned}$$

This completes the proof. □

The m.g.f. of $X \sim Hyp(a,n,N)$, althouh exists (since S_X is finite), can not be expressed in closed form.

12.1.4. The Poisson Distribution

Some event E (say number of cars crossing a particular bridge/tunnel) is occuring randomly over a period of time. Let X denotes the number of times E has occured in an unit interval (say $(0, 1]$).

To model probability distribution of X, partition the unit interval into a large number (say n where $n \to \infty$) of infinitesimal subintervals $(\frac{i-1}{n}, \frac{i}{n}]$, $i = 1, 2, \ldots, n$ of length $\frac{1}{n}$ each. In many situations, it may be relevant to assume that

(i) For each infinitesimal interval $(\frac{i-1}{n}, \frac{i}{n}]$, $i = 1, 2, \ldots, n$, the probability that E will occur in this interval is p_n and that it will not occur in this interval is $1 - p_n$; here $p_n \to 0$ as $n \to \infty$ and $np_n \to \lambda \in (0, \infty)$ as $n \to \infty$.

(ii) Chance of two or more occurences of E in any infinitesimal interval $(\frac{i-1}{n}, \frac{i}{n}]$, $i = 1, 2, \ldots, n$, is so small that it can be neglected.

(iii) occurrences of E in two disjoint infinitesimal intervals are independent.

$X \equiv X_n =$ the number of times event E occurs in $(0, 1] \sim Bin(n, p_n)$. The p.m.f. of X_n is

$$\begin{aligned}
f_n(k) &= \binom{n}{k} p_n^k (1-p_n)^{n-k} I_{\{0,1,\ldots,n\}}(k) \\
&= \frac{1}{k!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)(np_n)^k\left(1-\frac{np_n}{n}\right)^{n-k} I_{\{0,1,\ldots,n\}}(k) \\
&\to \frac{e^{-\lambda}\lambda^k}{k!} I_{\{0,1,\ldots,n\}}(k) \\
&= \begin{cases} \dfrac{e^{-\lambda}\lambda^k}{k!}, & k \in \{0,1,2,\ldots\}, \\ 0, & \text{otherwise.} \end{cases} \to \text{Poisson distribution } (Po(\lambda) : \lambda > 0) (\text{family of probability distributions}).
\end{aligned}$$

A r.v. X is said to have a Poisson distribution with parameter $\lambda > 0$ (written as $X \sim Po(\lambda)$) if its p.m.f. is given by

$$f_X(k) = P(X = k) = \begin{cases} \dfrac{e^{-\lambda}\lambda^k}{k!}, & k = 0, 1, 2, \ldots, \\ 0, & \text{otherwise.} \end{cases}$$

Theorem 12.10 (Poisson Approximation to Binomial Distribution)**.** *Let $X_n \sim Bin(n, p_n)$, $n = 1, 2, \ldots,$ where $p_n \in (0, 1)$, $n = 1, 2, \ldots$ and $\lim_{n\to\infty}(np_n) = \lambda$, for some $\lambda > 0$. Then*

$$\lim_{n\to\infty} f_{X_n}(k) = \lim_{n\to\infty} P(X_n = k) = \begin{cases} \dfrac{e^{-\lambda}\lambda^k}{k!}, & k = \{0, 1, 2, \ldots\}, \\ 0, & \textit{otherwise.} \end{cases}$$

Proof. As above. □

Remark 12.11. *If n is large and p is small ($p_n \to 0$ as $n \to \infty$) so that np is a fixed quantity in $(0, \infty)$ ($np_n \to \lambda > 0$) then Poisson distribution provides a good approximation to Binomial distribution.*

Example 12.12. *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 atleast two games.*

Solution: Let X denote the number of wins (successes) in 2500 games played by person.

Clearly $X \sim Bin(2500, 0.002)$, where $n = 2500$ and $np = 5$ (=λ, say) is fixed. Therefore,

$$P(X \geq 2) \approx P(Y \geq 2), \quad \text{where } Y \sim Po(5).$$

Thus, $P(X \geq 2) \approx 1 - (P(Y=0) + P(Y=1)) = 1 - (e^{-5} + 5e^{-5}) = 0.9596$.

Suppose that $X \sim Po(\lambda)$, for some $\lambda > 0$. Then for $r \in \{1, 2, \dots\}$, we have

$$\begin{aligned} E(X_{(r)}) &= E(X(X-1)\cdots(X-r+1)) \\ &= \sum_{k=0}^{\infty} k(k-1)(k-2)\cdots(k-r+1)\frac{e^{-\lambda}\lambda^k}{k!} \\ &= e^{-\lambda}\sum_{k=r}^{\infty}\frac{\lambda^k}{(k-r)!} = e^{-\lambda}\sum_{j=0}^{\infty}\frac{\lambda^{j+r}}{j!} = \lambda^r e^{-\lambda}\sum_{j=0}^{\infty}\frac{\lambda^j}{j!} = \lambda^r. \end{aligned}$$

Thus,

$$\begin{aligned} \mu_1 &= E(X) = E(X_{(1)}) = \lambda, \\ E(X^2) &= E(X_{(2)}) + E(X) = \lambda^2 + \lambda, \quad \mathrm{Var}(X) = E(X^2) - (E(X))^2 = \lambda \ (\sigma^2 = \mu_2) \quad \text{(Mean=Variance)}, \\ \mu_3' &= E(X^3) = \lambda(\lambda^2 + 3\lambda + 1), \\ \mu_4' &= E(X^4) = \lambda(\lambda^3 + 6\lambda^2 + 7\lambda + 1), \ \ \mu_3 = \lambda; \ \mu_4 = \lambda(3\lambda+1), \\ \beta_1 &= \frac{\mu_3}{\sigma^3} = \frac{1}{\sqrt{\lambda}}, \ \ \nu_2 = \nu_1 - 3 = \frac{\lambda(3\lambda+1)}{\lambda^2} - 3 = \frac{1}{\lambda}, \\ M_X(t) &= E(e^{tX}) = \sum_{k=0}^{\infty} e^{tk}\frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{(\lambda e^{tk})^k}{k!} = e^{-\lambda}e^{\lambda e^t} = e^{\lambda(e^t-1)}, \ \ t \in \mathbb{R}, \\ \psi_X(t) &= \ln M_X(t) = \lambda(e^t - 1), \quad \psi_X^r(t) = \lambda e^t, \ \ r = 1, 2, \dots, \\ &\Longrightarrow E(X) = \psi_X^1(0) = \lambda, \ \ \mathrm{Var}(X) = \psi_X^2(0) = \lambda. \end{aligned}$$

Theorem 12.13. *Let $X_1, X_2, \dots, X_k$ be independent r.v.'s such that $X_i \sim Po(\lambda_i)$, for some $\lambda_i > 0$, $i = 1, 2, \dots, k$. Then $Y = \sum_{i=1}^k X_i \sim Po(\lambda)$, where $\lambda = \sum_{i=1}^k \lambda_i$.*

Proof. For $t \in \mathbb{R}$,

$$\begin{aligned} M_Y(t) = E(e^{tY}) = E(e^{t\sum_{i=1}^k X_i}) = E\left(\prod_{i=1}^k e^{tX_i}\right) &= \prod_{i=1}^k E(e^{tX_i}), \quad \text{(independent of } X_i's) \\ &= \prod_{i=1}^k M_{X_i}(t) = \prod_{i=1}^k e^{\lambda_i(e^t-1)} = e^{\lambda(e^t-1)}. \end{aligned}$$

This implies that $Y \sim Po(\lambda)$, where $\lambda = \sum_{i=1}^k \lambda_i$. $\square$