CHAPTER 9

HYPOGEOMETRIC DISTRIBUTION – A DISCRETE LIFE TIME MODEL

9.1 INTRODUCTION

Numerous parametric models are used in the analysis of lifetime data and in the study of system reliability. Among univariate models, a few particular distributions occupy a central role because of their usefulness in a wide range of situations. Some of the important models in this category are the exponential, Weibull, gamma, lognormal. Shooman (1968) has examined aging and failure process of such life time models. Using probabilistic and/or statistical techniques analytical models of the reliability of a system can be derived. A system may be collections of equipments arrange to perform a function. As for e.g. CPU, missile, TV etc. The prediction of system reliability is based on the failure time distribution used. The equipments arranged in the system may be in series, parallel, standby or m out of n mode. A system with three components functions when at least two equipments function. Such a system is known as 2 out of 3 system or Triple Modular Redundant (TMR) system.


Consider a system with two statistically independent components having different failure rate with an exponential failure law, assuming a two-component stand by system with perfect switching equipment. Then the lifetime distribution of the standby system would be a two-stage hypoexponential distribution. Several properties and applications of such hypoexponential lifetime model are discussed by Trivedi (2002).

In this chapter, we have suggested modified version of geometric lifetime model discussed in earlier chapters, so called hypogeometric distribution. An advantage is that the hypogeometric lifetime model provides higher system
reliability rather than the system having geometric lifetime. This is demonstrated in section: 9.2. Some other properties of the model are derived and a simulated example is given to illustrate the method of estimation of the parameters.

9.2 DERIVATION AND PROPERTIES OF THE MODEL

Consider a system with two statistically independent components; parallel redundancy can be used in which both the components are initially operative simultaneously. Alternatively, we could initially keep the spare component in a power off state (de energized) and later upon the failure of the operative component, replace it by the spare. Assuming that a de energized component does not fail and that the failure detection and switching equipment is perfect. We can characterize the failure trial \( X \) of such a system in terms of the failure trials \( X_1 \) and \( X_2 \) of individual components by \( X = X_1 + X_2 \). Such a system is known to possess standby redundancy and the distribution of failure trials \( (X) \) of a system is called two stage hypogeometric distribution provided distribution of \( X \) is geometric with parameter \( q_i \) and the probability mass function (pmf)

\[
P(X_i = x_i) = p_i q_i^{x_i}; x_i = 0, 1, 2...; 0 < q_i < 1, p_i = 1 - q_i, i = 1, 2.
\]

Here \( p_x = P(X = x) = \sum_{x_i=0}^{x} P(X_1 = x_1)P(X_2 = x - x_1|X_1 = x_1) \]

\[
= \sum_{x_i=0}^{x} p_1 q_1^{x_i} p_2 q_2^{x-x_i}
\]

\[
= \frac{p_1 q_2}{q_2 - q_1} p_2 q_2^{x} + \frac{p_2 q_1}{q_1 - q_2} p_1 q_1^{x};
\]

\[
0 < q_i < 1, p_i = 1 - q_i, i = 1, 2.
\]  \hspace{1cm} (9.2.1)

which is the pmf of hypo geometric distribution .

Its cumulative probability distribution is given by

\[
F_X (x) = P(X \leq x)
\]

\[
= 1 - \frac{p_1 q_1}{q_1 - q_2} q_2^{x+1} - \frac{p_2 q_2}{q_2 - q_1} q_1^{x+1}
\]  \hspace{1cm} (9.2.2)

138
Laplace transform of hypogeeometric distribution:

Laplace transform of geometric variable $X_i$ is given by

$$L_{X_i}(S) = E(e^{-SX_i})$$

$$= \frac{p_i e^s}{e^s - q_i}.$$ 

Hence Laplace transform of hypogeometric variate $X = X_1 + X_2$ becomes

$$L_X(S) = L_{X_1}(S)L_{X_2}(S)$$

$$= \left( \frac{p_1 e^s}{e^s - q_1} \right) \left( \frac{p_2 e^s}{e^s - q_2} \right).$$

We expand this expression into a partial fraction as,

$$L_X(S) = c_1 \frac{p_1 e^s}{e^s - q_1} + c_2 \frac{p_2 e^s}{e^s - q_2}$$

where $c_1$ and $c_2$ are constant function of parameters $q_1$ and $q_2$.

On simplification, we find that

$$L_X(S) = \sum_{i=1}^{2} c_i \frac{p_i e^s}{e^s - q_i} \quad (9.2.3)$$

where $c_i = \prod_{j \neq i}^2 \frac{p_j q_i}{q_i - q_j} \ ; \ i = 1, 2.$

**Theorem:** Let $X_i$ 's ( $i = 1,2,\ldots,n$) are mutually independent and geometrically distributed random variables with parameter $q_i$ ( $q_i \neq q_j \ , \ i \neq j$) then $X = \sum_{i=1}^{n} X_i$ is an n-stage hypogeometric random variable having pmf

$$P(X = x) = \sum_{i=1}^{n} c_i p_i q_i^x$$

where $c_i = \prod_{j \neq i}^{n} \left( \frac{p_j q_i}{q_i - q_j} \right).$
**Proof:** Using result (9.2.3) and the technique of partial fraction expansion (see: Kobayashi(1978)) it can be shown that the Laplace transform of \( X \) can be written as

\[
L_x(s) = \sum_{i=1}^{n} c_i \frac{p_i e^s}{e^s - q_i}
\]

where \( c_i = \prod_{j=1}^{n} \left( \frac{p_j q_i}{q_i - q_j} \right) \).

Again from uniqueness theorem of Laplace transform it follows that the pmf of a random variable \( X \) is

\[
P(X = x) = \sum_{i=1}^{n} c_i p_i q_i^x
\]

which is denoted by

\( X \sim \text{Hypgeo}(q_1, q_2, \ldots, q_n) \).

**Corollary:** If \( X_1 \sim \text{Hypgeo}(q_1, q_2, \ldots, q_n) \) and \( X_2 \sim \text{Hypgeo}(q_{r+1}, q_{r+2}, \ldots, q_n) \) and \( X_1 \) and \( X_2 \) are independent then

\( X_1 + X_2 \sim \text{Hypgeo}(q_1, q_2, \ldots, q_r, q_{r+1}, q_{r+2}, \ldots, q_n) \).

The proof is obvious from the theorem.

**Cumulants of the model:**

The moment generating function of a hypogeometric variate \( X \) is given by

\[
M_X(t) = M_{X_1}(t)M_{X_1}(t)
\]

\[
= \prod_{i=1}^{2} \left( \frac{p_i}{1 - q_i e^t} \right).
\]

Hence, cumulant generating function becomes

\[
K_X(t) = \log M_X(t)
\]

\[
= \sum_{i=1}^{2} \left[ \log p_i - \log(1 - q_i e^t) \right].
\]

Now \( \frac{\partial^r K_X(t)}{\partial t^r} = \sum_{i=1}^{2} \left( \sum_{j=1}^{\infty} q_j^j j^{r-1} e^{jt} \right) \).
Hence, $K_r(x) = r^{th}$ cumulant of a random variable $X$

$$= \sum_{i=1}^{2} \left( \sum_{j=1}^{\infty} q_i^j j^{r-1} \right)$$

$$= \sum_{i=1}^{2} (K_r(x_i))$$

where $K_r(x_i)$ is the $r^{th}$ cumulant of a random variable $X_i$, $i = 1, 2$.

Using the recurrence relation for cumulants of geometric distribution the
recurrence relation for cumulants of hypogeometric distribution is given by

$$K_r(x) = \sum_{i=1}^{2} \left( q_i \frac{dK_r(x_i)}{dq_i} \right). \quad (9.2.4)$$

**Factorial cumulant:**

The probability generating function (pgf) of the hypogeometric distribution is
given by

$$G(z) = E(z^X) = \prod_{i=1}^{2} \left( \frac{p_i}{1 - q_i z} \right). \quad (9.2.5)$$

Hence

$$\log G(1 + z) = -\log \left(1 - \frac{q_1 z}{p_1} \right) - \log \left(1 - \frac{q_2 z}{p_2} \right)$$

Assuming $0 < \frac{q_1 z}{p_1} < 1$ and using the expansion of the function, $-\log(1 + x)$ we get,

$$\log G(1 + z) = \sum_{r=1}^{\infty} \left( \frac{q_1 z}{p_1} \right)^r + \sum_{r=1}^{\infty} \left( \frac{q_2 z}{p_2} \right)^r$$

Comparing the coefficient of $\frac{z^r}{r!}$ on both the sides we get

$$K_{(r)} = \text{the } r^{th} \text{ descending factorial cumulant}$$

$$= (r - 1)! \left[ \left( \frac{q_1}{p_1} \right)^r \right] + \left( \frac{q_2}{p_2} \right)^r; r = 1, 2, \ldots \quad (9.2.6)$$
Here $K_{(r)} = \text{coefficient of } \frac{r^r}{r!} \text{ in the expansion of } \log H_x(t)$, where $H_x(t)$ is descending factorial moment generating function given by $H_x(t) = E[(1+t)^x]$.

**Recurrence relations for probability and factorial moments:**

From the pgf defined in (9.2.5) we have,

$$G'(z) = \frac{dG(z)}{dz} = G(z) \left[ \frac{q_1}{1-q_1z} + \frac{q_2}{1-q_2z} \right]$$

Again differentiating r times with respect to $Z$ on both sides we get

$$G^{(r+1)}(z) = \frac{d^{r+1}G(z)}{dz^{r+1}}$$

$$= \sum_{j=0}^{r} \binom{r}{j} G^{(r-j)}(z) j! \left[ q_1^{j+1}(1-q_1z)^{-j-1} + q_2^{j+1}(1-q_2z)^{-j-1} \right]$$

(9.2.7)

Now consider the following two cases:

(i) Substituting $z = 0$ in the above equation- (9.2.7) we get the recurrence relation for probability of the model as,

$$pr+1 = \frac{\sum_{j=0}^{r} (q_1^{j+1} + q_2^{j+1}) pr-j}{r+1}; \quad r = 0, 1, 2, ...$$

(9.2.8)

where $p_{r+1} = P(X = r+1)$ as given in (9.2.1).

(ii) Substituting $z = 1$ in equation- (9.2.7) we get recurrence relation for descending factorial moments as,

$$\mu_{(r+1)} = \sum_{j=0}^{r} \binom{r}{j} \left[ \left( \frac{q_1}{p_1} \right)^{j+1} + \left( \frac{q_2}{p_2} \right)^{j+1} \right], \quad r = 0, 1, 2, ... \quad \text{with } \mu_{(0)} = 0.$$  

(9.2.9)

where $\mu_{r+1} = E[X(X-1)...(X-r+1)]$. 

142
Reliability of the model at trial $t$:

From equation (9.2.2) the reliability of the model is given by

\[ R_x(t) = 1 - F_x(t) \]

\[ = \frac{p_2q_1}{q_1-q_2}q_1^{t+1} + \frac{p_1q_2}{q_2-q_1}q_2^{t+1} \]  \hspace{1cm} (9.2.10)

9.3 APPLICATION OF THE MODEL

Consider the triple modular redundant (TMR) system in which the system with three components that require two or more components to function for the correct operation of the system. The reliability of such system for identical component is given by

\[ R_{TMR}(t) = 3R^2(t) - 2R^3(t) \]  \hspace{1cm} (9.3.1)

Where \( R(t) = q^{t+1} \)  \hspace{1cm} (9.3.2)

= the reliability of a system with single component and geometric life time model with parameter $q$, $0 < q < 1$.

Here

\[ R_{TMR}(t) \leq R(t) \text{ for } 1 \leq t \leq t_0 \]

\[ \geq R(t) \text{ for } t_0 \leq t \leq \infty \]

where $t_0$ is the solution of

\[ R_{TMR}(t) = R(t) \]

i.e. \[ 2R^3(t) - 3R^2(t) + R(t) = 0. \]

An improvement over this scheme is known as TMR $|_1$ simplex. In which not only the failed component but also one of the good component is discarded.

We consider the TMR $|_1$ simplex system on the line of Trivedi(2002) in the following way:
Let $X_1, X_2, X_3$ denote the number of trials to failure of the three components. In addition, let $W$ denotes the residual trials to failure of the selected components. Then $L = \min\{ X_1, X_2, X_3 \} + W$ denotes the trials to failure of TMR simplex system.

Let $X_i$'s are mutually independent identical and geometrically distributed number of trials to failure of the $i^{th}$ component, $i = 1, 2, 3$ with parameter $q, 0 < q < 1$. In addition, the distribution of $\min\{X_1, X_2, X_3\}$ is geometric with parameter $q^3$. Then the distribution of $L$ would be two-stage hypogeometric distribution with parameters $q_1 = q^3$ and $q_2 = q$.

The reliability expression of TMR simplex system can be obtained using (9.2.10) as

$$R(t)_{\text{TMR simplex}} = \frac{(1-q^3)R(t) - (1-q)q^2 R^3(t)}{1-q^2} \quad (9.3.3)$$

where $R(t) = q^{*1}$

The following table shows the comparison of the reliability of a system having single component($R(t)$), TMR system($R_{\text{TMR}}(t)$) and TMR simplex system $R(t)_{\text{TMR simplex}}$ for different choice of parameter $q$. 

144
<table>
<thead>
<tr>
<th>q</th>
<th>t</th>
<th>$R(t)$</th>
<th>$R_{TMR}(t)$</th>
<th>$R(t)_{TMR/simplex}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>1</td>
<td>0.01</td>
<td>0.000298</td>
<td>0.0100909</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.001</td>
<td>2.998E-06</td>
<td>1.009091E-03</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.0001</td>
<td>2.9998E-08</td>
<td>1.009091E-04</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.00001</td>
<td>2.99999E-10</td>
<td>1.009091E-05</td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td>0.25</td>
<td>0.15625</td>
<td>0.2890625</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.125</td>
<td>4.296875E-02</td>
<td>0.1455078</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.0625</td>
<td>1.123047E-02</td>
<td>7.28759E-02</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.03125</td>
<td>2.868652E-03</td>
<td>3.645325E-02</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.015625</td>
<td>7.247925E-04</td>
<td>1.822853E-02</td>
</tr>
<tr>
<td>0.8</td>
<td>1</td>
<td>0.64</td>
<td>0.7045121</td>
<td>0.7743489</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.512</td>
<td>0.5179966</td>
<td>0.6463226</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.4096</td>
<td>0.3658776</td>
<td>0.530802</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.32764</td>
<td>0.2517538</td>
<td>0.4316783</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.262144</td>
<td>0.1701297</td>
<td>0.3489457</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.2097152</td>
<td>0.1134947</td>
<td>0.2810012</td>
</tr>
<tr>
<td>0.95</td>
<td>1</td>
<td>0.9025</td>
<td>0.973335</td>
<td>0.9799799</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.857375</td>
<td>0.9447768</td>
<td>0.9624934</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.8145062</td>
<td>0.9095411</td>
<td>0.9413867</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.7737809</td>
<td>0.8696283</td>
<td>0.9174819</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.7350919</td>
<td>0.8266516</td>
<td>0.8914685</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.6983373</td>
<td>0.7819017</td>
<td>0.8639232</td>
</tr>
<tr>
<td>0.99</td>
<td>1</td>
<td>0.9801</td>
<td>0.9988278</td>
<td>0.9991208</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.970299</td>
<td>0.9974059</td>
<td>0.9982647</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.960596</td>
<td>0.9954643</td>
<td>0.997146</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.9509901</td>
<td>0.9930295</td>
<td>0.995375</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.9414802</td>
<td>0.9901271</td>
<td>0.9941624</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.9320654</td>
<td>0.9867817</td>
<td>0.9923181</td>
</tr>
</tbody>
</table>

The above comparison shows that hypogeometric lifetime model improves the reliability of the system.
9.4 ESTIMATION OF THE PARAMETERS
AND FITTING OF THE MODEL

From the result- (9.2.6) we have

\[ K_{(1)} = \frac{q_1}{p_1} + \frac{q_2}{p_2} \]

\[ K_{(2)} = \left(\frac{q_1}{p_1}\right)^2 + \left(\frac{q_2}{p_2}\right)^2. \]

Also,

\[ 2K_{(2)} - K_{(1)}^2 = \left(\frac{q_1}{p_1} - \frac{q_2}{p_2}\right)^2 \geq 0 \]

\[ \therefore 2K_{(2)} \geq K_{(1)}^2. \]

Hence, we can write

\[ q_2^2(K - 2K_{(1)} - 2) - 2q_2(K - K_{(1)}) + K = 0 \]  \hspace{1cm} (9.4.1)

where \( K = K_{(2)} - K_{(1)}^2 \).  \hspace{1cm} (9.4.2)

Solving (9.4.1) for \( q_2 \) we get

\[ q_2 = \frac{(K - K_{(1)}) \pm \sqrt{2K_{(2)} - K_{(1)}^2}}{K - 2K_{(1)} - 2}; 2K_{(2)} \geq K_{(1)}^2. \]  \hspace{1cm} (9.4.3)

Due to symmetry in \( K_{(1)} \) and \( K_{(2)} \) the two solutions for \( q_2 \) would be solution for \( q_1 \) and \( q_2 \).

Upon replacing \( K_{(1)} \) and \( K_{(2)} \) by simple factorial cumulants we can find moment estimators \( \hat{q}_1 \) and \( \hat{q}_2 \) of \( q_1 \) and \( q_2 \) respectively.

To illustrate the estimation procedure we consider the following simulated data, ( \( q_1 = 0.6 \), \( q_2 = 0.7 \)).
Substituting the results in (9.4.3), we get

\[ q_2 = 0.7274256 \quad \text{or} \quad 0.5415528 \]

i.e. \( q_2 = 0.7274256 \) and \( q_1 = 0.5415528 \)

Or

\[ q_2 = 0.5415528 \quad \text{and} \quad q_1 = 0.7274256 \]

Using these moment estimators in result (9.2.8) we can calculate probability for \( x \) and multiplying these probabilities by \( N = \sum f \) we get expected frequencies. Then by using test of goodness of fit one can test it for the given data.

<table>
<thead>
<tr>
<th>( x )</th>
<th>( f )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>1</td>
<td>19</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>3</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
</tr>
</tbody>
</table>

Here \( \hat{K}_1 = \hat{m}_1 = \frac{\sum f x}{\sum f} = 3.85 \)

\[ \hat{K}_2 = \frac{\sum f x^2}{\sum f} - m_1^2 = 12.3675 \]

Now \( \hat{K}_{(2)} = \hat{K}_2 - \hat{K}_1 = 8.5175 \)