CHAPTER IV

BI-LOGNORMAL
ROC CURVE ANALYSIS
4.1 Univariate Bi-Lognormal ROC curve analysis

4.1.1 Introduction

Normal distribution is assumed to describe the accuracy of biomarker/diagnostic test in the context of ROC curve analysis in most of the situations (Green and Swets (1966), Metz et al. (1998), Zhou et al. (2002)). However, many biomarker show skewed or asymmetrical distribution. Such skewed distribution closely fits to an important life distribution called lognormal distribution. Moreover, Bi-Normal model produces degenerate ROC curve when the sample size is small. In such a case, the use of Bi-Lognormal ROC model is extremely useful.

Lognormal distribution is a most widely used distribution for positively skewed data sets. It is characterized by log-transformed variable, using the location parameter \( \mu \) and shape parameter \( \sigma^2 \). It is symmetrical at the log level. The PDF of Lognormal distribution is given by Aitchison and Brown (1957). Let ‘Z’ be a random variable which is said to follow a lognormal distribution if the density function takes the following form

\[
f(z) = \frac{1}{z\sigma \sqrt{2\pi}} \exp\left[-\frac{1}{2} \left(\frac{\log z - \mu}{\sigma}\right)^2\right], \quad -\infty < z < \infty, \quad -\infty < \mu < \infty, \quad \sigma^2 > 0
\]  

and the CDF of \( Z \) is defined by

\[
F(z) = \Phi\left(\frac{\log z - \mu}{\sigma}\right)
\]  

where \( \Phi(.) \) is used to represent the CDF of standard normal variate.
This Chapter deals with the analysis of ROC curve and its AUC developed from Log-Normal distribution. If the data fit the lognormal distribution for both the populations viz. healthy and diseased, we recommend the use of Bi-Lognormal ROC model to fit a smooth ROC curve for evaluating the accuracy of biomarker. The proposed model is validated using simulated as well real life examples.

In Section 4.1.2, Bi-Lognormal ROC model is discussed. Section 4.1.3 discusses the properties of the Bi-Lognormal ROC model. AUC of Bi-Lognormal ROC curve and its asymptotic confidence interval are discussed in 4.1.4. Section 4.1.5 provides the computation of optimal cut-off point for classification using the proposed procedure. Section 4.1.6, illustrates the proposed procedure using simulated data set and CSF IgG index.

The research work of this Section has been published in Amala and Pundir (2012).

4.1.2 Bi-Lognormal ROC model

Let X and Y be two independent and log-normally distributed random variables with parameters \((\mu_x, \sigma_x^2)\) and \((\mu_y, \sigma_y^2)\) respectively with \(\mu_y > \mu_x\). Notationally, \(X \sim LN(\mu_x, \sigma_x^2)\) and \(Y \sim LN(\mu_y, \sigma_y^2)\). The ROC model developed under this assumption is named as Bi-Lognormal ROC model.

The CDF of X and Y are given as follows

\[
F(x) = \Phi \left( \frac{\log t - \mu_x}{\sigma_x} \right)
\]  

(4.3)
and $G(y) = \Phi\left(\frac{\log t - \mu_y}{\sigma_y}\right)$. \hspace{1cm} (4.4)

The FPR of Bi-lognormal ROC curve at the threshold ‘t’ is found to be

$$\bar{F}_X(t) = \Phi\left(\frac{\mu_x - \log t}{\sigma_x}\right).$$ \hspace{1cm} (4.5)

From (4.5), we can get the expression for threshold ‘t’ as

$$t = \exp\left\{\mu_x + \sigma_x \Phi^{-1}(1 - \bar{F}_X(t))\right\}. \hspace{1cm} (4.6)$$

The TPR of Bi-lognormal ROC curve at the threshold ‘t’ is found to be

$$\bar{G}_Y(t) = \Phi\left(\frac{\mu_y - \log t}{\sigma_y}\right).$$ \hspace{1cm} (4.7)

Since we know that the ROC model is TPR as a function of FPR, Bi-Lognormal ROC model is derived as

$$\text{ROC}(t) = \Phi\left[\frac{\mu_y - \mu_x}{\sigma_y} - \frac{\sigma_x}{\sigma_y} \Phi^{-1}[1 - \bar{F}_X(t)]\right], 0 \leq \bar{F}_X(t) \leq 1 \hspace{1cm} (4.8)$$

$$= \Phi\left[a - b\Phi^{-1}[1 - \bar{F}_X(t)]\right]. \hspace{1cm} (4.9)$$

Alternatively, Bi-Lognormal ROC model can be written as

$$\text{ROC}(t) = \Phi\left[\frac{\mu_y - \mu_x}{\sigma_y} + \frac{\sigma_x}{\sigma_y} \Phi^{-1}[\bar{F}_X(t)]\right], 0 \leq \bar{F}_X(t) \leq 1 \hspace{1cm} (4.10)$$

$$= \Phi\left[a + b\Phi^{-1}[\bar{F}_X(t)]\right] \hspace{1cm} (4.11)$$
where \( a = \frac{\mu_y - \mu_x}{\sigma_y} \) and \( b = \frac{\sigma_x}{\sigma_y} \).

Bi-Lognormal ROC curve can be obtained by plotting \( \overline{F}_X(t) \) on X-axis and ROC(t) on Y-axis after substituting the parametric estimates from the sample. It can also be estimated by directly plotting \( \overline{F}_X(t) \) on X-axis and \( \overline{G}_Y(t) \) on Y-axis.

The parameters can be estimated from the sample data by any of the standard estimation procedures. For example, the MLE of the parameters of Lognormal distribution given by Shen (1998) are as follows

\[
\hat{\mu}_x = \exp \left\{ \bar{x} + \frac{s_x^2}{2} \right\},
\]

\[
\hat{\mu}_y = \exp \left\{ \bar{y} + \frac{s_y^2}{2} \right\},
\]

\[
\hat{\sigma}_x^2 = \exp \left\{ 2\bar{x} + s_x^2 \right\} \left( e^{s_x^2} - 1 \right)
\]

and \( \hat{\sigma}_y^2 = \exp \left\{ 2\bar{y} + s_y^2 \right\} \left( e^{s_y^2} - 1 \right) \)

where \( s_x^2 = \frac{\sum (x_i - \bar{x})^2}{m} \) and \( s_y^2 = \frac{\sum (y_i - \bar{y})^2}{n} \).

Substituting the above estimates of parameters in (4.5) and (4.7), one can get the estimates of \( \overline{F}_X(t) \) and \( \overline{G}_Y(t) \).
4.1.3 Properties

Once the ROC curve is plotted, it is important to study some intrinsic properties in order to highlight some key points regarding the accuracy of biomarker. The proposed Bi-Lognormal model satisfies the basic properties of the conventional Bi-Normal or NP ROC curve. The properties are

1. The Bi-Lognormal ROC curve is monotonically increasing in the positive quadrant lying between 0 and 1.

**Proof:** Let

\[
\text{ROC}(t) = \Phi \left( \frac{\mu_y - \mu_x}{\sigma_y} - \frac{\sigma_x}{\sigma_y} \Phi^{-1}[1 - F_X(t)] \right)
\]

be a continuous function. Differentiating \( \text{ROC}(t) \) with respect to \( F_X(t) \), we get

\[
\text{ROC}'(t) = \frac{d\text{ROC}(t)}{dF_X(t)} = \frac{b\phi \left(a - b\Phi^{-1}(1 - F_X(t))\right)}{\Phi \left(\Phi^{-1}(1 - F_X(t))\right)} > 0
\]

(4.14)

where \( \phi(.) \) and \( \Phi^{-1}(.) \) are the PDF and inverse CDF of standard normal variate respectively. From (4.14), it is obvious that the proposed model is monotonically increasing function in \([0, 1]\) which is one of the desirable property of the ROC curve.

2. The Bi-Lognormal ROC curve is a concave in nature.

**Proof:**

For a function \( \text{ROC}(t) \) to be a concave function, the second order derivative of \( \text{ROC}(t) \) with respect to \( F_X(t) \) should be negative.
Differentiating (4.14) with respect to $\bar{F}_X(t)$, we get

$$\text{ROC}'(t) = \frac{\phi\left[\Phi^{-1}(1 - \bar{F}_X(t))\right] \frac{d}{d\bar{F}_X(t)} \left[ \text{b}\phi\left[a - b\Phi^{-1}(1 - \bar{F}_X(t))\right]\right] - \text{b}\phi\left[a - b\Phi^{-1}(1 - \bar{F}_X(t))\right] \frac{d}{d\bar{F}_X(t)} \left[\phi\left[\Phi^{-1}(1 - \bar{F}_X(t))\right]\right]}{\left[\phi\left[\Phi^{-1}(1 - \bar{F}_X(t))\right]\right]^2}$$

(4.15)

where

$$D_1 = \frac{d}{d\bar{F}_X(t)} \left[ \text{b}\phi\left[a - b\Phi^{-1}(1 - \bar{F}_X(t))\right]\right]$$

and

$$D_2 = \frac{d}{d\bar{F}_X(t)} \left[\phi\left[\Phi^{-1}(1 - \bar{F}_X(t))\right]\right].$$

Now on evaluating $D_1$, we get

$$D_1 = \frac{d}{d\bar{F}_X(t)} \left[ \frac{\text{b}}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2} \left[a - b\Phi^{-1}(1 - \bar{F}_X(t))\right]^2\right\}\right]$$

$$= -b^2 \left[a - b\Phi^{-1}(1 - \bar{F}_X(t))\right] \frac{\phi\left[a - b\Phi^{-1}(1 - \bar{F}_X(t))\right]}{\phi\left[\Phi^{-1}(1 - \bar{F}_X(t))\right]}$$

(4.16)

and

$$D_2 = \Phi^{-1}(1 - \bar{F}_X(t)).$$

(4.17)
Substituting (4.16) and (4.17) in (4.15), we get

\[
\text{ROC}'(t) = -b\Phi\left[a - b\Phi^{-1}\left(1 - F_X(t)\right)\right]\Phi^{-1}\left(1 - F_X(t)\right) + \\
\frac{\Phi^{-1}\left(1 - F_X(t)\right)}{\phi\left[\Phi^{-1}\left(1 - F_X(t)\right)\right]^2} < 0. 
\] (4.18)

Hence, Bi-Lognormal ROC curve is concave in nature is proved.

3. Bi-Lognormal ROC curve is asymmetric.

The Bi-Lognormal ROC curve satisfies the asymmetry property analogous to Bi-Normal ROC curve as discussed by Hughes and Bhattacharya (2013). It possess the same analytical form for K-L(f, g) and K-L(g, f) as in Bi-Normal case which are given as follows:

\[
\text{KL}(f, g) = \frac{1}{2}\left[\frac{\sigma^2}{\sigma^2_y} + \frac{(\hat{\mu}_x - \hat{\mu}_y)^2}{\sigma^2_y} + \ln\left(\frac{\sigma^2}{\sigma^2_y}\right)\right] 
\] (4.19)

\[
\text{KL}(g, f) = \frac{1}{2}\left[\frac{\sigma^2}{\sigma^2_x} + \frac{(\hat{\mu}_x - \hat{\mu}_y)^2}{\sigma^2_x} + \ln\left(\frac{\sigma^2}{\sigma^2_x}\right)\right] 
\] (4.20)

where the values of the estimated parameters \(\hat{\mu}_x, \hat{\mu}_y, \sigma_x^2\) and \(\sigma_y^2\) can be obtained from (4.12) and (4.13). The Bi-Lognormal ROC curve is TPR asymmetric when \(\frac{\sigma_x}{\sigma_y} < 1\) and TNR asymmetric when \(\frac{\sigma_x}{\sigma_y} > 1\).

4. The slope of Bi-Lognormal ROC curve at any operating point ‘t’ is given by

\[
\text{slope} = \frac{\sigma_x}{\sigma_y} \exp\left\{-\frac{1}{2}\left[\left(\frac{\ln - \hat{\mu}_x}{\sigma^2_y}\right)^2 - \left(\frac{\ln - \hat{\mu}_x}{\sigma^2_x}\right)^2\right]\right\}. 
\] (4.21)
4.1.4 AUC of Bi-Lognormal ROC curve and its asymptotic confidence interval

For the estimation of AUC, we adopt the logarithmic transformation to the random variable \( X \) and \( Y \) in order to make the computations easy. We know that if \( X \) and \( Y \) follows log-normal distribution then logarithm of \( X \) and \( Y \) follows normal distribution. Since it is easier to work with normal assumption as compared to lognormal, we transform \( x = \ln(X) \) and \( y = \ln(Y) \); here ‘\( \ln \)’ stands for natural logarithm (\( \log_e \)).

\[ \text{AUC} = P(Y > X) = P(\ln Y > \ln X) = P(y > x). \quad (4.22) \]

We know that

\[ y - x \sim N(\mu_y - \mu_x, \sigma_x^2 + \sigma_y^2). \]

Therefore,

\[ P(y - x > 0) = P \left( Z > \frac{\mu_x - \mu_y}{\sqrt{\sigma_y^2 + \sigma_x^2}} \right) = 1 - P \left( Z \leq \frac{\mu_x - \mu_y}{\sqrt{\sigma_y^2 + \sigma_x^2}} \right) = 1 - \Phi \left( \frac{- (\mu_y - \mu_x)}{\sqrt{\sigma_y^2 + \sigma_x^2}} \right), \]

Hence,

\[ \text{AUC} = \Phi \left( \frac{\mu_y - \mu_x}{\sqrt{\sigma_y^2 + \sigma_x^2}} \right) = \Phi(\delta) \quad (4.23) \]

where \( \delta = \frac{\mu_y - \mu_x}{\sqrt{\sigma_y^2 + \sigma_x^2}} \quad (4.24) \)
By substituting (4.12) and (4.13) in (4.22), we get the estimate of accuracy as

\[
\hat{AUC} = \Phi \left( \frac{\hat{\mu}_y - \hat{\mu}_x}{\sqrt{\hat{\sigma}_y^2 + \hat{\sigma}_x^2}} \right).
\]

(4.25)

The MLE of \( \delta \) is given by

\[
\hat{\delta} = \frac{\hat{\mu}_y - \hat{\mu}_x}{\sqrt{\hat{\sigma}_x^2 + \hat{\sigma}_y^2}}.
\]

(4.26)

Since \( \Phi(\hat{\delta}) \) is a monotonically increasing function of \( \hat{\delta} \), it is enough to find the variance and standard error of \( \hat{\delta} \) for determining the confidence interval for AUC. Since \( \delta \) is a function of parameters \( \theta = (\mu_x, \mu_y, \sigma_x^2, \sigma_y^2) \), we will adopt the delta method to find the approximate variance and standard error of \( \hat{\delta} \).

Therefore, by definition

\[
V(\hat{\delta}) = \left( \frac{\partial \hat{\delta}}{\partial \hat{\mu}_y} \right)^2 V(\hat{\mu}_y) + \left( \frac{\partial \hat{\delta}}{\partial \hat{\mu}_x} \right)^2 V(\hat{\mu}_x) + \left( \frac{\partial \hat{\delta}}{\partial \hat{\sigma}_y^2} \right)^2 V(\hat{\sigma}_y^2) + \left( \frac{\partial \hat{\delta}}{\partial \hat{\sigma}_x^2} \right)^2 V(\hat{\sigma}_x^2) + 2 \left( \frac{\partial \hat{\delta}}{\partial \hat{\mu}_y} \right) \left( \frac{\partial \hat{\delta}}{\partial \hat{\mu}_x} \right) \text{Cov}(\hat{\mu}_y, \hat{\mu}_x) + 2 \left( \frac{\partial \hat{\delta}}{\partial \hat{\sigma}_y^2} \right) \left( \frac{\partial \hat{\delta}}{\partial \hat{\sigma}_x^2} \right) \text{cov}(\hat{\sigma}_y^2, \hat{\sigma}_x^2)
\]

\[
= \left( \frac{1}{\sqrt{\hat{\sigma}_y^2 + \hat{\sigma}_x^2}} \right)^2 \frac{\sigma_x^2}{m} + \left( \frac{-1}{\sqrt{\hat{\sigma}_y^2 + \hat{\sigma}_x^2}} \right)^2 \frac{\sigma_y^2}{n} + \left( \frac{-(\hat{\mu}_y - \hat{\mu}_x)}{2(\hat{\sigma}_y^2 + \hat{\sigma}_x^2)} \right)^2 \frac{2\sigma_x^4}{m-1} + \left( \frac{-\hat{\mu}_y - \hat{\mu}_x)}{2(\hat{\sigma}_y^2 + \hat{\sigma}_x^2)} \right)^2 \frac{2\sigma_y^4}{n-1}
\]

\[
= \frac{1}{(\hat{\sigma}_y^2 + \hat{\sigma}_x^2)} \left( \frac{\sigma_y^2}{n} + \frac{\sigma_x^2}{m} + \frac{(\hat{\mu}_y - \hat{\mu}_x)^2}{2(\hat{\sigma}_y^2 + \hat{\sigma}_x^2)} \left[ \frac{\sigma_y^4}{(n-1)} + \frac{\sigma_x^4}{(m-1)} \right] \right). \quad (4.27)
\]
\( V(\hat{\delta}) \) can be obtained by substituting the estimated values from (4.12) and (4.13) in (4.27). Then the calculation of confidence interval for \( \hat{\text{AUC}} \) is straightforward and it is given as follows

\[
\Phi \left[ \hat{\delta} - Z_{\alpha} \sqrt{V(\hat{\delta})}, \hat{\delta} + Z_{\alpha} \sqrt{V(\hat{\delta})} \right]
\]

(4.28)

where \( \alpha \) is the level of significance and \( z_{\alpha/2} \) is the critical value of Z for a two tailed test at level of significance \( \alpha \).

### 4.1.5 Determination of optimal cut-off

In addition with the assessment of the biomarker/diagnostic test, the optimal cut-off value for future classification is needed. Fluss et al. (2005) suggested a methodology to find the optimal cut-off point using Youden index. The method is obtaining the maximum of the difference between \( F(t) \) and \( G(t) \) where \( F \) and \( G \) represents the CDF of healthy and diseased individuals respectively. Explicitly,

\[
t^* = \max_t [F(t) - G(t)] = \max_t \left\{ \Phi \left[ \frac{\ln t - \mu_x}{\sigma_x} \right] - \Phi \left[ \frac{\ln t - \mu_y}{\sigma_y} \right] \right\}.
\]

(4.29)

\( 't^* ' \) will yield the optimal cut-off point by differentiating \( \Phi \left[ \frac{\ln t - \mu_x}{\sigma_x} \right] - \Phi \left[ \frac{\ln t - \mu_y}{\sigma_y} \right] \) with respect to ‘t’ and equating it to zero.
Chapter IV: Bi-Lognormal ROC Curve Analysis

It has been obtained as

\[
    t^* = \exp \left[ \frac{(\mu_y \sigma_x^2 - \mu_x \sigma_y^2) - \sigma_x \sigma_y \sqrt{(\mu_y - \mu_x)^2 + 2(\sigma_x^2 - \sigma_y^2) \ln \left( \frac{\sigma_x}{\sigma_y} \right)}}{\sigma_x^2 - \sigma_y^2} \right].
\]

(4.30)

The value of \( t^* \) can be obtained by substituting the estimated values from (4.12) and (4.13) in (4.30).

4.1.6 Numerical example

(a) Simulation studies

In this Section, we have illustrated the proposed Bi-Lognormal ROC model using simulated data set. The data has been generated from Lognormal distribution by assuming various parametric values viz., \( \mu_x = \{3, 4, 5.5, 5.7\} \), \( \mu_y = \{8, 7, 7, 6\} \), \( \sigma_x = \{1, 1.2, 1.5, 0.5\} \), \( \sigma_y = \{2, 2.3, 1.2, 0.6\} \). The estimated parameters, AUC and 95% asymptotic confidence interval for AUC is presented in Table 4.1. The Bi-Lognormal ROC curves for different values of parameters are shown in Fig. 4.1.

<table>
<thead>
<tr>
<th>( \hat{\mu}_x )</th>
<th>( \hat{\mu}_y )</th>
<th>( \hat{\sigma}_x )</th>
<th>( \hat{\sigma}_y )</th>
<th>AUC</th>
<th>95% confidence interval for AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.777</td>
<td>8.478</td>
<td>0.993</td>
<td>1.95</td>
<td>0.995</td>
<td>[0.975, 0.999]</td>
</tr>
<tr>
<td>3.973</td>
<td>7.058</td>
<td>1.246</td>
<td>2.69</td>
<td>0.851</td>
<td>[0.732, 0.928]</td>
</tr>
<tr>
<td>5.496</td>
<td>6.864</td>
<td>1.718</td>
<td>1.20</td>
<td>0.743</td>
<td>[0.608, 0.849]</td>
</tr>
<tr>
<td>5.614</td>
<td>5.939</td>
<td>0.429</td>
<td>0.566</td>
<td>0.676</td>
<td>[0.536, 0.795]</td>
</tr>
</tbody>
</table>
Real life example

Multiple Sclerosis (MS) is an inflammatory disease in which the fatty myelin sheaths around the axons of brain and spinal cord are damaged, leading to demyelination and scarring as well as a broad spectrum of signs and symptoms. For detecting Multiple Sclerosis, one of the possible laboratory tests is the Cerebro Spinal Fluid (CSF) Immunoglobulin G (IgG) index. CSF IgG index is defined as the ratio of (IgG) in CSF /IgG in Serum to CSF-albumin/ Serum albumin. The suggested cut-off value for this biomarker is 0.73 mg/dL (Hische et al. 1982). Higher values of CSF IgG index are suspicious for multiple sclerosis. We consider the Multiple Sclerosis data (Zhou et al., 2002) to validate the Bi-Lognormal ROC model. The data set consisted of 40 subjects. Among them 20 are affected by MS disease and 20 are controls for MS but affected other neurological disorders. The results of GOF test for the data set has been given in Table 4.2. The GOF between the theoretical and empirical densities of healthy and diseased markers are plotted in Fig. 4.2 and 4.3.
Table 4.2 The results of GOF test of CSF IgG index for Lognormal distribution

<table>
<thead>
<tr>
<th>Test</th>
<th>Statistic</th>
<th>Rank</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>H</td>
<td>D</td>
<td></td>
</tr>
<tr>
<td>Kolmogorov-Smirnov</td>
<td>0.15833</td>
<td>0.20026</td>
<td>0.64127</td>
</tr>
<tr>
<td>Chi-Square</td>
<td>0.67215</td>
<td>1.0776</td>
<td>0.71457</td>
</tr>
<tr>
<td>Anderson Darling’s</td>
<td>0.65108</td>
<td>0.72213</td>
<td>-</td>
</tr>
</tbody>
</table>

Fig. 4.2 GOF test between the theoretical and empirical densities of healthy marker

Fig. 4.3 GOF test between the theoretical and empirical densities of diseased marker

The MLE of parameters are found to be $\mu_x = 0.643; \mu_y = 1.6655; \sigma_x = 0.2891$ and $\sigma_y = 0.8969$. The Bi-Lognormal ROC curve has showed an accuracy of 0.853 (85%) with an estimated confidence interval for AUC is [0.7089, 0.9393]. It is to be noted that, the sensitivity and the specificity of CSF IgG index using Bi-Lognormal ROC curve are 77.4 % and 76.8 % respectively for the cut-off point is 0.855 g/L. Table 4.3 presents the sensitivity and specificity of the biomarker for various cut-offs. One criterion for determining the optimal threshold of a test/biomarker is to minimize the difference between sensitivity and specificity with the constraint that Sn>Sp and hence 0.855 g/L can be taken as optimal cut-off from Table 4.3.
The Non-Parametric ROC curve for CSF IgG index is also plotted in order to make a comparison with Bi-Lognormal ROC curve. It has given an accuracy of 0.843 (84%) with an estimated confidence interval as [0.7174, 0.9676]. The sensitivity and the specificity of CSF IgG index using NP ROC curve are found to be 80% and 75% respectively at the threshold 0.855g/L. The Bi-Lognormal and NP ROC curves plotted for CSF IgG index is shown in Fig. 4.4. The smooth curve represents the Bi-Lognormal and the jagged curve represents the NP curve.

Fig. 4.4 Non-parametric and Bi-Lognormal ROC curve plotted for CSF IgG index
Table 4.3 Sensitivity and specificity analysis of CSF IgG index using Bi-Lognormal ROC curve

<table>
<thead>
<tr>
<th>t</th>
<th>FPR</th>
<th>Sensitivity (Sn)</th>
<th>Specificity (Sp)</th>
<th>Min</th>
<th>Sn – Sp</th>
<th>, Sn &gt; Sp</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.16</td>
<td>0.9887</td>
<td>0.9996</td>
<td>0.0113</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.285</td>
<td>0.8929</td>
<td>0.9926</td>
<td>0.1071</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>0.7363</td>
<td>0.9725</td>
<td>0.2637</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>0.7363</td>
<td>0.9725</td>
<td>0.2637</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.405</td>
<td>0.7290</td>
<td>0.9713</td>
<td>0.2710</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.425</td>
<td>0.6995</td>
<td>0.9661</td>
<td>0.3005</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.445</td>
<td>0.6701</td>
<td>0.9604</td>
<td>0.3299</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.455</td>
<td>0.6555</td>
<td>0.9574</td>
<td>0.3445</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.475</td>
<td>0.6266</td>
<td>0.9510</td>
<td>0.3734</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.505</td>
<td>0.5843</td>
<td>0.9407</td>
<td>0.4157</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.565</td>
<td>0.5043</td>
<td>0.9176</td>
<td>0.4957</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.625</td>
<td>0.4322</td>
<td>0.8914</td>
<td>0.5678</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.64</td>
<td>0.4155</td>
<td>0.8845</td>
<td>0.5845</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.65</td>
<td>0.4047</td>
<td>0.8798</td>
<td>0.5953</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.67</td>
<td>0.3837</td>
<td>0.8702</td>
<td>0.6163</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.685</td>
<td>0.3686</td>
<td>0.8629</td>
<td>0.6314</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.745</td>
<td>0.3133</td>
<td>0.8326</td>
<td>0.6867</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.825</td>
<td>0.2514</td>
<td>0.7904</td>
<td>0.7486</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>0.855</strong></td>
<td><strong>0.2314</strong></td>
<td><strong>0.7743</strong></td>
<td><strong>0.7686</strong></td>
<td></td>
<td></td>
<td><strong>0.0057</strong></td>
</tr>
<tr>
<td>0.86</td>
<td>0.2282</td>
<td>0.7716</td>
<td>0.7718</td>
<td></td>
<td></td>
<td>0.0002</td>
</tr>
<tr>
<td>0.87</td>
<td>0.2219</td>
<td>0.7662</td>
<td>0.7781</td>
<td></td>
<td></td>
<td>0.0119</td>
</tr>
<tr>
<td>0.89</td>
<td>0.2100</td>
<td>0.7554</td>
<td>0.7900</td>
<td></td>
<td></td>
<td>0.0346</td>
</tr>
<tr>
<td>0.955</td>
<td>0.1753</td>
<td>0.7202</td>
<td>0.8247</td>
<td></td>
<td></td>
<td>0.1045</td>
</tr>
<tr>
<td>1.015</td>
<td>0.1485</td>
<td>0.6879</td>
<td>0.8515</td>
<td></td>
<td></td>
<td>0.1636</td>
</tr>
<tr>
<td>1.03</td>
<td>0.1425</td>
<td>0.6799</td>
<td>0.8575</td>
<td></td>
<td></td>
<td>0.1776</td>
</tr>
<tr>
<td>1.075</td>
<td>0.1258</td>
<td>0.6561</td>
<td>0.8742</td>
<td></td>
<td></td>
<td>0.2181</td>
</tr>
<tr>
<td>1.14</td>
<td>0.1053</td>
<td>0.6224</td>
<td>0.8947</td>
<td></td>
<td></td>
<td>0.2723</td>
</tr>
<tr>
<td>1.265</td>
<td>0.0751</td>
<td>0.5605</td>
<td>0.9249</td>
<td></td>
<td></td>
<td>0.3644</td>
</tr>
<tr>
<td>1.5</td>
<td>0.0404</td>
<td>0.4565</td>
<td>0.9596</td>
<td></td>
<td></td>
<td>0.5031</td>
</tr>
<tr>
<td>1.795</td>
<td>0.0193</td>
<td>0.3502</td>
<td>0.9807</td>
<td></td>
<td></td>
<td>0.6305</td>
</tr>
<tr>
<td>1.995</td>
<td>0.0120</td>
<td>0.2922</td>
<td>0.9880</td>
<td></td>
<td></td>
<td>0.6958</td>
</tr>
<tr>
<td>2.05</td>
<td>0.0105</td>
<td>0.2781</td>
<td>0.9895</td>
<td></td>
<td></td>
<td>0.7114</td>
</tr>
<tr>
<td>2.11</td>
<td>0.0092</td>
<td>0.2634</td>
<td>0.9908</td>
<td></td>
<td></td>
<td>0.7274</td>
</tr>
<tr>
<td>2.165</td>
<td>0.0081</td>
<td>0.2507</td>
<td>0.9919</td>
<td></td>
<td></td>
<td>0.7412</td>
</tr>
<tr>
<td>2.33</td>
<td>0.0056</td>
<td>0.2162</td>
<td>0.9944</td>
<td></td>
<td></td>
<td>0.7782</td>
</tr>
<tr>
<td>2.61</td>
<td>0.0031</td>
<td>0.1687</td>
<td>0.9969</td>
<td></td>
<td></td>
<td>0.8282</td>
</tr>
<tr>
<td>2.735</td>
<td>0.0024</td>
<td>0.1513</td>
<td>0.9976</td>
<td></td>
<td></td>
<td>0.8463</td>
</tr>
<tr>
<td>2.815</td>
<td>0.0020</td>
<td>0.1412</td>
<td>0.9980</td>
<td></td>
<td></td>
<td>0.8568</td>
</tr>
<tr>
<td>3.06</td>
<td>0.0012</td>
<td>0.1145</td>
<td>0.9988</td>
<td></td>
<td></td>
<td>0.8843</td>
</tr>
<tr>
<td>4.23</td>
<td>0.0002</td>
<td>0.0446</td>
<td>0.9998</td>
<td></td>
<td></td>
<td>0.9552</td>
</tr>
</tbody>
</table>
Table 4.4 Sensitivity and specificity analysis of CSF IgG index using Non-Parametric method

<table>
<thead>
<tr>
<th>t</th>
<th>FPR</th>
<th>TPR (Sensitivity)</th>
<th>TNR (Specificity)</th>
<th>Min</th>
<th>(\text{Sn} - \text{Sp}), Sn &gt; Sp</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.85</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>0.16</td>
<td>0.95</td>
<td>1</td>
<td>0.05</td>
<td>0.95</td>
<td></td>
</tr>
<tr>
<td>0.285</td>
<td>0.9</td>
<td>1</td>
<td>0.1</td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td>0.405</td>
<td>0.75</td>
<td>1</td>
<td>0.25</td>
<td>0.75</td>
<td></td>
</tr>
<tr>
<td>0.445</td>
<td>0.7</td>
<td>0.95</td>
<td>0.3</td>
<td>0.65</td>
<td></td>
</tr>
<tr>
<td>0.475</td>
<td>0.65</td>
<td>0.9</td>
<td>0.35</td>
<td>0.55</td>
<td></td>
</tr>
<tr>
<td>0.505</td>
<td>0.6</td>
<td>0.9</td>
<td>0.4</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>0.565</td>
<td>0.6</td>
<td>0.85</td>
<td>0.4</td>
<td>0.45</td>
<td></td>
</tr>
<tr>
<td>0.64</td>
<td>0.55</td>
<td>0.85</td>
<td>0.45</td>
<td>0.4</td>
<td></td>
</tr>
<tr>
<td>0.65</td>
<td>0.5</td>
<td>0.8</td>
<td>0.5</td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td>0.67</td>
<td>0.45</td>
<td>0.8</td>
<td>0.55</td>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td>0.745</td>
<td>0.35</td>
<td>0.8</td>
<td>0.65</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td>0.825</td>
<td>0.3</td>
<td>0.8</td>
<td>0.7</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td><strong>0.855</strong></td>
<td>0.25</td>
<td><strong>0.8</strong></td>
<td><strong>0.75</strong></td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td><strong>0.86</strong></td>
<td>0.25</td>
<td><strong>0.8</strong></td>
<td><strong>0.75</strong></td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>0.87</td>
<td>0.2</td>
<td>0.75</td>
<td>0.8</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>0.89</td>
<td>0.15</td>
<td>0.75</td>
<td>0.85</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>0.955</td>
<td>0.15</td>
<td>0.7</td>
<td>0.85</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td>1.015</td>
<td>0.15</td>
<td>0.65</td>
<td>0.85</td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td>1.14</td>
<td>0.05</td>
<td>0.6</td>
<td>0.95</td>
<td>0.35</td>
<td></td>
</tr>
<tr>
<td>1.265</td>
<td>0</td>
<td>0.6</td>
<td>1</td>
<td>0.4</td>
<td></td>
</tr>
<tr>
<td>1.5</td>
<td>0</td>
<td>0.55</td>
<td>1</td>
<td>0.45</td>
<td></td>
</tr>
</tbody>
</table>

Though the accuracy of Bi-Lognormal and NP ROC curves seems to be closer, the width of confidence interval is narrow for Bi-Lognormal when compared to NP ROC curve which implies the standard error is lesser for Bi-Lognormal ROC curve.
4.2 Bi-Variate Lognormal ROC curve analysis

4.2.1. Introduction

In biomedical research, two or more biomarkers may be available for diagnosis of a particular disease. Selecting one single biomarker which ideally discriminates diseased group from healthy group is a confront in a diagnostic process. Frequently, most of the people use the accuracy measure, AUC to choose the best diagnostic biomarker among the available biomarkers for diagnosis. Some authors have tried to combine the multiple biomarkers by an optimal linear combination to increase the discriminatory power. In this Section, we propose an alternative method that combines two continuous biomarkers by direct bi-variate modeling of ROC curve under lognormality assumption. The proposed method is applied to simulated data set and prostate cancer diagnostic biomarkers.

The parametric ROC curve analysis for univariate biomarker has been extensively studied by many authors. Bi-Normal ROC curve was first studied by Green and Swets (1966), Bi-logistic ROC curve for rating data (Oglive and Creelman, 1968), Bi-Exponential ROC model (Betinec (2008), Pundir and Amala (2014: e)), Bi-Gamma ROC curve (Dorfman et al., 1996), Bi-Lomax ROC (Campbell and Ratnaparkhi, 1993) Generalized Bi-Exponential ROC curve by Hussian (2011), Bi-Lognormal ROC model (Amala and Pundir, 2012), Bi-Rayleigh ROC curve (Pundir and Amala, 2012: a, b), a proper Bi-Weibull ROC curve (Pundir and Amala, 2014: d) and a review of all parametric ROC models in case of continuous data (Pundir and Amala, 2014: a).

Most of the diagnostic procedures yield multiple parameters/outcomes/measurements during the process. For example, the multiple measurements may be
cholesterol level, gender, age and weight of the subject for Coronary and Vascular diseases. Here, the biomarker of interest is cholesterol level and gender, age and weight are the supporting variables called covariates which may give valuable information for the better classification. Sometimes the efficiency of biomarker may be affected by covariates and hence it is important to adjust for potential covariates. One could find the study of ROC curves with covariates (Pepe (2000), Schisterman et al. (2004)). Pepe and Thompson (2000) have provided a distribution-free linear combination of biomarkers with covariates to enhance the diagnostic accuracy. They have also done a comparative study with Su and Liu’s and Logistic regression methodologies (Su and Liu (1993), Richards et al. (1996)). Schisterman et al., (2004) have presented the ROC modeling to adjust for covariates using multivariate regression modeling techniques.

Apart from covariates, there may be two or more potential biomarkers available for the diagnosis of a specific disease. For example, let us consider Prostate cancer diagnostic biomarkers data sets. The Prostate-Specific Antigen (PSA) is an emblematic example: a case study from Etzioni et al. (1999). Two biomarkers namely total serum PSA (tPSA) and ratio of (percent) free to total PSA (fPSA) are significant in detecting the prostate cancer. In this case, one of the classical way to choose a best biomarker among the potential biomarkers is through the comparison of accuracy indices Area Under the ROC curve (AUC)’s of those biomarkers (Hanley and McNeil (1982 and 1983), Moladianovitch et al., 2006). But it is always believed that the use of all potential biomarkers simultaneously provides a better diagnosis than the use of single biomarker. Now, how to combine the information from multiple biomarkers to get relevant accuracy without losing any information? Su and Liu (1993) provided a best linear combination of biomarkers using
Fisher linear discriminant co-efficient with and without restrictions on the covariance matrix using bi-normality model. If X and Y denote the normally distributed bi-variate biomarkers in healthy and diseased group of size 'm' and 'n' respectively i.e. \( X \sim N(\mu_X, \Sigma_X) \) and \( Y \sim N(\mu_Y, \Sigma_Y) \), where \( \mu_X = (\mu_{x1}, \mu_{x2})' \), \( \mu_Y = (\mu_{y1}, \mu_{y2})' \) and \( \Sigma_X \) and \( \Sigma_Y \) are the variance-covariance matrix of X and Y respectively. Let 'a' be the best linear combination of biomarkers obtained from Fisher linear discriminant function, say

\[
a = (\Sigma_X + \Sigma_Y)^{-1}(\mu_Y - \mu_X)
\]

(4.31)

with the corresponding ROC model

\[
ROC(t) = \Phi \left\{ \frac{a^T(\mu_Y - \mu_X) + \Phi^{-1}[F_X(t)]\sqrt{a^T\Sigma_X a}}{\sqrt{a^T\Sigma_Y a}} \right\}, 0 \leq F_X(t) \leq 1
\]

(4.32)

where \( \Phi(.) \) is the CDF of standard normal distribution. The ROC model given in (4.32) possesses the following

\[
AUC = \Phi \left\{ (\mu_Y - \mu_X)^T(\Sigma_X + \Sigma_Y)^{-1}(\mu_Y - \mu_X) \right\}.
\]

(4.33)

Schisterman et al. (2004) have discussed the Su and Liu’s methodology combined with the covariate effect. Hsu and Hsueh (2013) and Hsu et al. (2014) have investigated a linear combination of biomarkers that maximizes the partial Area Under the ROC curve (pROC) over the range of FPRs. Also, when two biomarkers are identified from different biological or functional systems, it may not be appropriate to combine them using a linear combination. Hence, for this problem Wang and Liu (2012) explored the two biomarkers in bi-variate settings non-parametrically. In this Chapter, we have provided a simple and alternative parametric approach to combine information from two continuous biomarkers.
with emphasis on each biomarker to extract the overall accuracy of classification using bi-variate lognormal distribution. When multiple outcomes are available for a diagnostic process, if the distributions of multiple outcomes are violating the multivariate normal assumption, and then Logistic regression (Richards et al., 1996) can be used as an alternative to linear discriminant procedure (Pepe, 2000). Gupta et al. (2013) discussed the estimation of reliability parameter from a bi-variate log-normal data for the case of equal means and unequal means.

This Section is organized in the following way: In Section 4.2.2, ROC curve analysis for bi-variate biomarker using lognormal distribution has been introduced with its AUC. Section 4.2.3, discusses the procedure to find standard error of $\hat{AUC}$ using Monte Carlo (MC) simulation technique with its confidence interval. In Section 4.2.4, the proposed model has been validated using simulated data set as well as the real life application. The bi-variate normal ROC curve is also discussed using simulation studies.

The research work of this Section is submitted for publication.

4.2.2 Bi-Variate Bi-Lognormal ROC Curve Model

Sometimes two biomarkers may be highly associated with the presence of disease (Ma and Huang, 2007) and it is indispensable to consider both the biomarkers for detecting the disease. Suppose let $(X, Y)$ represents the biomarker values of healthy and diseased group. Let $(X_1, X_2)$ be two sets of related biomarkers taken from healthy group. Let $(Y_1, Y_2)$ be two sets of biomarkers taken from diseased group. Obviously, $(X_1, X_2)$ and $(Y_1, Y_2)$ are independent pairs of bi-variate biomarkers for healthy and diseased subjects respectively. A particular subject is identified as diseased when the values of $Y_1$ and $Y_2$
are large enough (Wang and Li, 2012). It is also to be noted that a subject is classified as diseased if both the related biomarkers are large enough.

The CDFs for healthy and diseased group are defined by
\[ F_X(t_1,t_2) = P(X_1 \leq t_1, X_2 \leq t_2) \] and \[ F_Y(t_1,t_2) = P(Y_1 \leq t_1, Y_2 \leq t_2) \] respectively. FPR and TPR in the bi-variate criteria can be defined by \[ P(X_1 > t_1, X_2 > t_2) \] and \[ P(Y_1 > t_1, Y_2 > t_2) \] respectively.

Let \( X \) and \( Y \) be random vectors of continuous biomarker values from the healthy and diseased groups respectively assumes bi-variate lognormal distribution with parameters \( \mu_x, \mu_y, \Sigma_x \) and \( \Sigma_y \) where \( \Sigma 's \) are of order \( 2 \times 2 \). Symbolically, \( X \sim LN(\mu_x, \Sigma_x) \) and \( Y \sim LN(\mu_y, \Sigma_y) \). where \( \mu_x = \begin{pmatrix} \mu_{x_1} \\ \mu_{x_2} \end{pmatrix} \), \( \mu_y = \begin{pmatrix} \mu_{y_1} \\ \mu_{y_2} \end{pmatrix} \),

\[
\Sigma_x = \begin{pmatrix} \sigma_{x_1}^2 & \rho_{x_1} \sigma_{x_1} \sigma_{x_2} \\ \rho_{x_1} \sigma_{x_2} \sigma_{x_1} & \sigma_{x_2}^2 \end{pmatrix} \quad \text{and} \quad \Sigma_y = \begin{pmatrix} \sigma_{y_1}^2 & \rho_{y_1} \sigma_{y_1} \sigma_{y_2} \\ \rho_{y_1} \sigma_{y_2} \sigma_{y_1} & \sigma_{y_2}^2 \end{pmatrix}.
\]

Then the PDF of \( X \) defined by Aitchison and Brown (1957) takes the following form

\[
f(x_1, x_2) = \frac{1}{2\pi \sigma_{x_1} \sigma_{x_2} \sqrt{1-\rho_x^2}} \exp \left\{ -\frac{1}{2(1-\rho_x^2)} \left( \frac{\ln x_1 - \mu_{x_1}}{\sigma_{x_1}} \right)^2 + \left( \frac{\ln x_2 - \mu_{x_2}}{\sigma_{x_2}} \right)^2 \right\} \left( 1 - 2\rho_x \left( \frac{\ln x_1 - \mu_{x_1}}{\sigma_{x_1}} \right) \left( \frac{\ln x_2 - \mu_{x_2}}{\sigma_{x_2}} \right) \right). \]

(4.35)

Let \( Z_1 = \ln(X_1) \) and \( Z_2 = \ln(X_2) \), then \( Z_1 \sim N\left(\mu_{x_1}, \sigma_{x_1}^2\right) \), \( Z_2 \sim N\left(\mu_{x_2}, \sigma_{x_2}^2\right) \) where \( E[Z_j] = \mu_{zj} \) and \( V[Z_j] = \sigma_{zj}^2 \) \((j=1, 2)\) and the correlation between \( X_1 \) and \( X_2 \) is \( \rho_x \). Similarly, assume that
Chapter IV: Bi-Lognormal ROC Curve Analysis

\[ Y_1 \sim LN(\mu_{y_1}, \sigma_{y_1}^2) \quad \text{and} \quad Y_2 \sim LN(\mu_{y_2}, \sigma_{y_2}^2) \quad \text{and} \quad Y_1 \text{ and } Y_2 \text{ are related then the joint density of } Y_1 \text{ and } Y_2 \text{ is called bi-variate lognormal distribution, i.e. } Y \sim LN(\mu_y, \Sigma_y). \]

The PDF of \( Y \) is given by

\[
f(y_1, y_2) = \frac{1}{2\pi \sigma_{y_1} \sigma_{y_2} \sqrt{1-\rho_y^2}} \exp \left\{ \frac{-1}{2(1-\rho_y^2)} \left[ \left( \frac{\ln y_1 - \mu_{y_1}}{\sigma_{y_1}} \right)^2 + \left( \frac{\ln y_2 - \mu_{y_2}}{\sigma_{y_2}} \right)^2 \right] - 2\rho_y \left( \frac{\ln y_1 - \mu_{y_1}}{\sigma_{y_1}} \right) \left( \frac{\ln y_2 - \mu_{y_2}}{\sigma_{y_2}} \right) \right\}. \tag{4.36}
\]

Let \( W_1 = \ln(Y_1) \) and \( W_2 = \ln(Y_2) \), then \( W_1 \sim N(\mu_{y_1}, \sigma_{y_1}^2) \), \( W_2 \sim N(\mu_{y_2}, \sigma_{y_2}^2) \) where \( E[Z_j] = \mu_{z_j} \) and \( V[Z_j] = \sigma_{z_j}^2 \) (j=1, 2) and the correlation between \( Y_1 \) and \( Y_2 \) is \( \rho_y \). The correlation coefficients \( \rho_x \) and \( \rho_y \) for healthy and diseased can be estimated from the following expressions given by Johnson and Kotz (1972) as

\[
\rho_x = \frac{\exp(\rho_x \sigma_{x_1} \sigma_{x_2}) - 1}{\sqrt{\exp(\sigma_{x_1}^2) - 1} \sqrt{\exp(\sigma_{x_2}^2) - 1}} \quad \text{and} \quad \rho_y = \frac{\exp(\rho_w \sigma_{y_1} \sigma_{y_2}) - 1}{\sqrt{\exp(\sigma_{y_1}^2) - 1} \sqrt{\exp(\sigma_{y_2}^2) - 1}} \tag{4.37}
\]

where \( \rho_z \) and \( \rho_w \) are the correlation co-efficients between \( (X_1, X_2) \) and \( (Y_1, Y_2) \) respectively.

The explicit form of FPR and TPR in case of bi-variate setting is analytically...
complicated. The FPR from bi-variate biomarkers at the thresholds \( t_1 \) and \( t_2 \) is given as

\[
FPR(t_1, t_2) = \psi \left( \frac{\ln(t_2) - \mu_{x_2}}{\sigma_{x_2}} \right)
\]

where

\[
\psi(c) = \int_{c}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2}} \Phi \left( \frac{(\ln(t_1) - \mu_{x_1} - \rho_{x,y} \sigma_{x_1} \nu)}{\sqrt{1-\rho_{x,y}^2} \sigma_{x_1}} \right) dv.
\]

The TPR from a bi-variate biomarker at the thresholds \( t_1 \) and \( t_2 \) are derived as

\[
TPR(t_1, t_2) = \xi \left( \frac{\ln(t_2) - \mu_{y_2}}{\sigma_{y_2}} \right)
\]

where

\[
\xi(c) = \int_{c}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2}} \Phi \left( \frac{(\ln(t_1) - \mu_{y_1} - \rho_{x,y} \sigma_{y_1} \nu)}{\sqrt{1-\rho_{x,y}^2} \sigma_{y_1}} \right) dv.
\]

The TPR and FPR are estimated by substituting the maximum likelihood estimates of parameters in (4.38) and (4.39). The proposed method do not uses any optimality strategy or any special assumption; it just follows the methodology as the univariate ROC curve analysis do. The same strategy can be extended to a bi-variate normal distribution by replacing \( \ln(t_1) \) and \( \ln(t_2) \) by just \( t_1 \) and \( t_2 \) without any difficulty. It is also to be noted that the bi-variate lognormal ROC model is one of the special case of bi-variate Normal ROC model using the fact that if \( Z \sim MVLN(\mu, \Sigma) \) then \( \log(Z) \sim MVN(\mu, \Sigma) \), where ‘MVLN’ and ‘MVN’ stands for ‘Multi Variate Log-Normal’ and ‘Multi Variate Normal’ distribution.
respectively. In this situation, the results of bi-variate Lognormal ROC procedure applied to Z and the results of bi-variate Normal ROC curve applied to log(Z) is one and the same.

The AUC of bi-variate Bi-Normal ROC curve takes the following form

\[
AUC = P(Y_1 > X_1, Y_2 > X_2) = \int \int I(y_1 > x_1, y_2 > x_2) dF_X(t_1, t_2) dF_Y(t_1, t_2). \tag{4.40}
\]

The AUC cannot be expressed in an assessable closed form. One would estimate the accuracy by adopting any of the numerical procedure such as trapezoidal, Simpson’s approximation, etc. A special function that calculate the AUC of bi-variate Lognormal ROC curve is presented in Appendix I. In the following Section, the approximate variance of \( \hat{AUC} \) using Monte Carlo simulation is obtained.

4.2.3 Approximate variance of \( \hat{AUC} \) using Monte Carlo simulation technique

When it is impossible to compute the exact result with the deterministic procedure, the estimates can be computed approximately by simulation techniques like Bootstrap, Jackknife procedure, Monte Carlo simulation procedure, etc. Since a closed form expression for AUC under bi-variate Bi-lognormal ROC curve is not possible, we obtain the standard error of AUC estimate by Monte Carlo simulation technique.

Monte Carlo methods are computational algorithm done by repeated random sampling for large number of times to compute the estimates Anderson (1986). Monte Carlo method follows the following steps:
Step 1: Determine the parametric model that fit best to the chosen data. Once the model has been selected, extract the estimates of parameters related to the selected model.

Step 2: By taking the above estimates as the parametric values, generate Monte Carlo sample from bi-variate lognormal distribution of size m for healthy and n for diseased group (i.e.) \( X_i^* = \begin{pmatrix} X_{i1}^* \\ X_{i2}^* \end{pmatrix} \) and \( Y_j^* = \begin{pmatrix} Y_{1j}^* \\ Y_{2j}^* \end{pmatrix} \); \( i=1, 2 \ldots m; j=1, 2 \ldots n; \) and based on the sample generated compute AUC, say \( \hat{AUC} \), using (4.40).

Step 3: Repeat Step 2 for large number of times, say 10000.

Step 4: The standard deviations of 10000 computed AUCs is nothing but the Monte Carlo estimate of standard error say \( \hat{Se}(\hat{AUC}) \).

4.2.4 Numerical example

(i) Simulation studies

(a) Bi-Variate Lognormal ROC model

We know that, when \( \zeta \sim LN(\mu, \Sigma) \) then \( \ln \zeta \sim N(\mu, \Sigma) \). To generate \( \zeta \), generate \( \ln \zeta \) with mean and variance equal to the mean and variance of \( \ln \zeta \) then take the exponent of the resulting values. Using this fact, the data has been generated from bi-variate lognormal distribution of size \( (m, n) = (50, 50) \) with same parametric values and for different values of correlation co-efficients. The assumed correlation coefficients, estimated parameters and
correlation co-efficients, \( \hat{\text{AUC}} \), \( \text{Se}(\hat{\text{AUC}}) \) and 95\% Monte Carlo confidence interval for AUC of bi-variate Lognormal ROC curve are presented in Table 4.5.

Table 4.5 Estimated parameters, \( \hat{\text{AUC}} \), \( \text{Se}(\hat{\text{AUC}}) \) and 95\% Monte Carlo confidence interval for AUC of bi-variate Lognormal ROC curve for different correlation co-efficients

<table>
<thead>
<tr>
<th>( \rho_x = \rho_y )</th>
<th>( \hat{\mu} ) and ( \hat{\Sigma} )</th>
<th>( \hat{\rho}_x )</th>
<th>( \hat{\rho}_y )</th>
<th>( \hat{\text{AUC}}(\text{Se}(\hat{\text{AUC}})) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9</td>
<td>( \hat{\mu}_x = (13.72, 5.093) ), ( \hat{\mu}_y = (91.27, 58.21) ), ( \hat{\Sigma}_x = (49.11, 16.8, 16.8), \hat{\Sigma}_y = (3383, 2212, 2212) )</td>
<td>0.85</td>
<td>0.90</td>
<td>0.9937 (0.0048) [0.9843, 1.0000]</td>
</tr>
<tr>
<td>0.7</td>
<td>( \hat{\mu}_x = (13.99, 4.9) ), ( \hat{\mu}_y = (73.86, 43.29) ), ( \hat{\Sigma}_x = (40.86, 7.14, 7.14), \hat{\Sigma}_y = (410, 440.7, 1179) )</td>
<td>0.60</td>
<td>0.63</td>
<td>0.9993 (0.0035) [0.9923, 1.0000]</td>
</tr>
<tr>
<td>0.4</td>
<td>( \hat{\mu}_x = (13.05, 4.51) ), ( \hat{\mu}_y = (88.04, 56.38) ), ( \hat{\Sigma}_x = (37.6, 5.13, 5.13), \hat{\Sigma}_y = (3017, 657.3, 657.3) )</td>
<td>0.34</td>
<td>0.39</td>
<td>0.9986 (0.0026) [0.9935, 1.0000]</td>
</tr>
<tr>
<td>0.2</td>
<td>( \hat{\mu}_x = (14.04, 4.64) ), ( \hat{\mu}_y = (74.84, 60.65) ), ( \hat{\Sigma}_x = (45.39, 4.43, 4.43), \hat{\Sigma}_y = (1495, 60.22, 60.22) )</td>
<td>0.25</td>
<td>0.034</td>
<td>0.9985 (0.0019) [0.9948, 1.0000]</td>
</tr>
</tbody>
</table>

From Table 4.5, we observe that the AUC is insensitive to the strength of correlation. But the standard error of AUC decreases as the correlation decreases, which mean that the measure of correlation has an impact on the error component associated with the accuracy. We also observe that the accuracy increases as the difference between mean of two groups increases with decrease in standard error of AUC.
Then, the data has been generated from bi-variate lognormal distribution of size \((m, n) = (50, 50)\) for different values of parameters with fixed correlation coefficient \((\rho_x = \rho_y = 0.7)\), to study the behavior of bi-variate Lognormal ROC curve and its AUC. The estimated parameters, \(\hat{AUC}, \text{Se}(\hat{AUC})\) and 95% Monte Carlo confidence interval for AUC of bi-variate Lognormal ROC curve are presented in Table 4.6.

**Table 4.6 Estimated parameters, \(\hat{AUC}, \text{Se}(\hat{AUC})\) and 95% Monte Carlo confidence interval for AUC of bi-variate Lognormal ROC curve for different parametric values with \(\rho_x = \rho_y = 0.7\)**

<table>
<thead>
<tr>
<th>Estimated parameters bi-variate Lognormal distribution</th>
<th>(\hat{AUC}(\text{Se}(\hat{AUC})))</th>
<th>95% confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\hat{\mu}_x = \begin{pmatrix} 14.33 \ 5.01 \end{pmatrix}, \hat{\mu}_y = \begin{pmatrix} 15.46 \ 6.73 \end{pmatrix} ), (\hat{\Sigma}_x = \begin{pmatrix} 31.90 &amp; 9.95 \ 9.95 &amp; 6.6 \end{pmatrix}, \hat{\Sigma}_y = \begin{pmatrix} 34.3 &amp; 13.27 \ 13.27 &amp; 9.94 \end{pmatrix} )</td>
<td>0.637 ( (0.0666) )</td>
<td>[0.5065, 0.7675]</td>
</tr>
<tr>
<td>(\hat{\mu}_x = \begin{pmatrix} 13.41 \ 5.028 \end{pmatrix}, \hat{\mu}_y = \begin{pmatrix} 18.51 \ 7.28 \end{pmatrix} ), (\hat{\Sigma}_x = \begin{pmatrix} 29.30 &amp; 10.27 \ 10.27 &amp; 7.94 \end{pmatrix}, \hat{\Sigma}_y = \begin{pmatrix} 52.24 &amp; 10.6 \ 10.6 &amp; 6.79 \end{pmatrix} )</td>
<td>0.760 ( (0.0613) )</td>
<td>[0.6399, 0.8801]</td>
</tr>
<tr>
<td>(\hat{\mu}_x = \begin{pmatrix} 12.49 \ 5.105 \end{pmatrix}, \hat{\mu}_y = \begin{pmatrix} 22.047 \ 8.869 \end{pmatrix} ), (\hat{\Sigma}_x = \begin{pmatrix} 28.62 &amp; 11.06 \ 11.06 &amp; 9.52 \end{pmatrix}, \hat{\Sigma}_y = \begin{pmatrix} 70.17 &amp; 27.93 \ 27.93 &amp; 17.29 \end{pmatrix} )</td>
<td>0.859 ( (0.0508) )</td>
<td>[0.7594, 0.9586]</td>
</tr>
<tr>
<td>(\hat{\mu}_x = \begin{pmatrix} 14.57 \ 4.87 \end{pmatrix}, \hat{\mu}_y = \begin{pmatrix} 64.51 \ 7.79 \end{pmatrix} ), (\hat{\Sigma}_x = \begin{pmatrix} 35.74 &amp; 11.5 \ 11.5 &amp; 5.86 \end{pmatrix}, \hat{\Sigma}_y = \begin{pmatrix} 527.61 &amp; 45.29 \ 45.29 &amp; 8.25 \end{pmatrix} )</td>
<td>0.939 ( (0.0277) )</td>
<td>[0.8847, 0.9933]</td>
</tr>
</tbody>
</table>

From Table 4.6 we observe that, as the estimated parametric values deviate more among each other especially \(\mu_i\)s, the AUC increases with decrease in standard error.
(b) Bi-Variate Normal ROC model

In this Section, the data has been generated from bi-variate normal distribution of size \((m, n) = (50, 50)\) with same parametric values and for different values of correlation co-efficients. The assumed correlation coefficients, estimated parameters and correlation co-efficients, \(\hat{AUC}, \text{Se}(\hat{AUC})\) and 95% Monte Carlo confidence interval for AUC of bi-variate normal ROC curve are presented in Table 4.7.

### Table 4.7 Estimated parameters, \(\hat{AUC}, \text{Se}(\hat{AUC})\) and 95% Monte Carlo confidence interval for AUC of bi-variate Normal ROC curve for different correlation co-efficients

<table>
<thead>
<tr>
<th>(\rho_x) ((\hat{\rho}_x))</th>
<th>(\rho_y) ((\hat{\rho}_y))</th>
<th>(\hat{\mu}) and (\hat{\Sigma})</th>
<th>(\hat{AUC}(\text{Se}(\hat{AUC}))) 95% confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2 ((0.12))</td>
<td>0.21</td>
<td>(\hat{\mu}_x = \begin{pmatrix} 2.46 \ 1.38 \end{pmatrix}, \hat{\mu}_y = \begin{pmatrix} 3.6 \ 2.9 \end{pmatrix} , \hat{\Sigma}_x = \begin{pmatrix} 0.195 &amp; 0.027 \ 0.027 &amp; 0.261 \end{pmatrix}, \hat{\Sigma}_y = \begin{pmatrix} 0.325 &amp; 0.059 \ 0.059 &amp; 0.242 \end{pmatrix} )</td>
<td>0.979 ((0.0197)) ([0.9404, 1.000])</td>
</tr>
<tr>
<td>0.4 ((0.44))</td>
<td>0.38</td>
<td>(\hat{\mu}_x = \begin{pmatrix} 2.47 \ 1.41 \end{pmatrix}, \hat{\mu}_y = \begin{pmatrix} 3.49 \ 2.88 \end{pmatrix} , \hat{\Sigma}_x = \begin{pmatrix} 0.1925 &amp; 0.1015 \ 0.1015 &amp; 0.278 \end{pmatrix}, \hat{\Sigma}_y = \begin{pmatrix} 0.2307 &amp; 1.089 \ 1.089 &amp; 3.489 \end{pmatrix} )</td>
<td>0.9659 ((0.0214)) ([0.924, 1.000])</td>
</tr>
<tr>
<td>0.7 ((0.73))</td>
<td>0.83</td>
<td>(\hat{\mu}_x = \begin{pmatrix} 2.423 \ 1.495 \end{pmatrix}, \hat{\mu}_y = \begin{pmatrix} 3.66 \ 2.92 \end{pmatrix} , \hat{\Sigma}_x = \begin{pmatrix} 0.299 &amp; 0.610 \ 0.610 &amp; 0.372 \end{pmatrix}, \hat{\Sigma}_y = \begin{pmatrix} 0.345 &amp; 0.331 \ 0.331 &amp; 0.455 \end{pmatrix} )</td>
<td>0.9494 ((0.0249)) ([0.9006, 0.998])</td>
</tr>
<tr>
<td>0.9 ((0.83))</td>
<td>0.93</td>
<td>(\hat{\mu}_x = \begin{pmatrix} 2.47 \ 1.47 \end{pmatrix}, \hat{\mu}_y = \begin{pmatrix} 3.49 \ 2.72 \end{pmatrix} , \hat{\Sigma}_x = \begin{pmatrix} 0.224 &amp; 0.193 \ 0.193 &amp; 0.243 \end{pmatrix}, \hat{\Sigma}_y = \begin{pmatrix} 0.364 &amp; 0.391 \ 0.391 &amp; 0.486 \end{pmatrix} )</td>
<td>0.926 ((0.0273)) ([0.8725, 0.9795])</td>
</tr>
</tbody>
</table>
From Table 4.7, we observe that the AUC is insensitive to the strength of correlation. But the standard error of AUC is decreases as the correlation decreases, which mean that the measure of correlation has an impact on the error component associated with the accuracy. We also observe that the accuracy increases as the difference between mean of two group increases with decrease in standard error of AUC. Now with fixed correlation coefficient and different values of parameters, the behavior of bi-variate normal ROC curve and its AUC is studied. The estimated parameters, \( \hat{\text{AUC}}, \hat{\text{Se(AUC)}} \) and 95% Monte Carlo confidence interval for AUC of bi-variate normal ROC curve are presented in Table 4.8.

<table>
<thead>
<tr>
<th>Estimated parameters bi-variate Normal distribution</th>
<th>( \hat{\text{AUC}}(\text{Se(AUC)}) ) 95% confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{\mu}_x = \left( \begin{array}{c} 2.479 \ 1.443 \end{array} \right) ), ( \hat{\mu}_y = \left( \begin{array}{c} 2.869 \ 2.358 \end{array} \right) ), ( \hat{\Sigma}_x = \left( \begin{array}{cc} 1.075 &amp; 0.29 \ 0.29 &amp; 0.165 \end{array} \right) ), ( \hat{\Sigma}_y = \left( \begin{array}{cc} 1.94 &amp; 1.309 \ 1.309 &amp; 1.942 \end{array} \right) )</td>
<td>( (0.62, 0.78) ) ( (0.0699, 0.0681) ) ( [0.483, 0.757] ) ( [0.647, 0.914] )</td>
</tr>
<tr>
<td>( \hat{\mu}_x = \left( \begin{array}{c} 2.410 \ 0.816 \end{array} \right) ), ( \hat{\mu}_y = \left( \begin{array}{c} 4.118 \ 1.740 \end{array} \right) ), ( \hat{\Sigma}_x = \left( \begin{array}{cc} 0.419 &amp; 0.144 \ 0.144 &amp; 0.141 \end{array} \right) ), ( \hat{\Sigma}_y = \left( \begin{array}{cc} 3.056 &amp; 1.094 \ 1.094 &amp; 0.891 \end{array} \right) )</td>
<td>( (0.78, 0.97) ) ( (0.0681, 0.0135) ) ( [0.483, 0.757] ) ( [0.647, 0.914] )</td>
</tr>
<tr>
<td>( \hat{\mu}_x = \left( \begin{array}{c} 2.47 \ 1.45 \end{array} \right) ), ( \hat{\mu}_y = \left( \begin{array}{c} 3.044 \ 2.055 \end{array} \right) ), ( \hat{\Sigma}_x = \left( \begin{array}{cc} 0.1613 &amp; 1.098 \ 1.098 &amp; 0.209 \end{array} \right) ), ( \hat{\Sigma}_y = \left( \begin{array}{cc} 0.149 &amp; 0.13 \ 0.13 &amp; 1.86 \end{array} \right) )</td>
<td>( (0.87, 0.97) ) ( (0.0426, 0.0135) ) ( [0.787, 0.954] ) ( [0.787, 0.997] )</td>
</tr>
<tr>
<td>( \hat{\mu}_x = \left( \begin{array}{c} 2.9 \ 0.91 \end{array} \right) ), ( \hat{\mu}_y = \left( \begin{array}{c} 8.26 \ 4.1 \end{array} \right) ), ( \hat{\Sigma}_x = \left( \begin{array}{cc} 0.845 &amp; 0.645 \ 0.645 &amp; 1.713 \end{array} \right) ), ( \hat{\Sigma}_y = \left( \begin{array}{cc} 4.48 &amp; 2.114 \ 2.114 &amp; 2.305 \end{array} \right) )</td>
<td>( (0.97, 1.00) ) ( (0.0135, 0.0135) ) ( [0.944, 0.997] ) ( [0.944, 0.997] )</td>
</tr>
</tbody>
</table>
From Table 4.8, we observe that the accuracy increases as the difference between mean of two group increases and the standard error value follow the decreasing pattern as the accuracy increases.

(ii) **Real life example**

In this Section, the proposed method is applied to Prostate cancer biomarkers given by Etzioni (1999). PSA is a biomarker which is significant in detecting the prostate cancer. The data consist of 50 randomly chosen individuals who are affected by prostate cancer and 50 healthy individuals who participated in a lung cancer prevention trial. The two correlated prostate cancer biomarkers namely total serum PSA and the ratio of (percent) free to total PSA are considered. Among these biomarkers tPSA have higher AUC than fPSA and hence it is preferred to assess the accuracy of diagnosis for prostate cancer.

If both the biomarkers are significant in identifying the disease, we recommend to use a bi-variate ROC modeling in order to use both the indispensable biomarker. The data has to be evaluated for GOF for bi-variate lognormal distribution. Now, how would one tests a set of data that have come from a multivariate normal or lognormal distribution? The multivariate normality can be tested using Mardia’s test by Henze (1994), Henze and Zirkler (1990), BHEP multivariate tests (Henze and Wagner (1990), Royston (1983)), etc. Here, the log transformed data has been tested for GOF of multivariate normality using Royston’s test, since if ln(z) is normal, then z would be lognormaly distributed automatically. The Royston’s statistic for X and Y are 2.4032 and 0.6679 for healthy and diseased group respectively with the corresponding p-values are 0.1557 and 0.6219. The Q-Q plots of GOF for H and D are presented in Fig. 4.5 and 4.6 respectively.
Once the GOF is evaluated, the ROC curve for bi-variate biomarkers is plotted to the data. The ROC curve obtained is shown in Fig. 4.4. The area under the bi-variate lognormal ROC curve is found to be 0.8572 (86%) with an estimated standard error 0.04476. If we plot the ROC curve for individual biomarker using univariate Bi-lognormal ROC curve the accuracies were found to be 0.84 (tPSA) and 0.88 (tPSA). The ROC curve for univariate biomarkers and bi-variate biomarkers are shown in Fig. 4.7, 4.8 and 4.9 respectively. By applying the procedure of Su and Liu, the estimate of AUC is found to be 0.89. By applying the empirical logistic regression technique as suggested by Pepe and Thompson (2000), the accuracy is found to be 0.89 with an estimated accuracy of 0.032. The ROC curve plotted for Su and Liu’s methodology is shown in Fig. 4.10.
Chapter IV: Bi-Lognormal ROC Curve Analysis

Fig. 4.7 Bi-Lognormal ROC curve for fPSA biomarker with AUC as 0.84

Fig. 4.8 Bi-Lognormal ROC curve for tPSA biomarker with AUC as 0.88

Fig. 4.9 Bi-variate Bi-Lognormal ROC curve with AUC as 0.8572

Fig. 4.10 Su and Liu’s ROC curve with AUC as 0.89
4.3 Conclusion

ROC curve analysis of biomarker based on lognormality assumption has been carried out in this Chapter. If the data is asymmetric or positively skewed then the usage of Bi-Normal ROC model will not yield an accurate AUC. Hence, the need for an alternative model is required viz. Bi-Lognormal ROC model. It was observed that Bi-Lognormal ROC curve is monotonically increasing, concave in nature and TPR asymmetric. The Bi-Lognormal model is applied to CSF IgG index and it is found that CSF IgG index is able to identify the disease with an accuracy of 85% with the sensitivity and specificity of CSF IgG index are 77.4% and 76.8% respectively.

In many diagnostic processes, multiple diagnostic biomarkers are recorded and there is a need to combine those biomarkers to predict the accuracy. We have proposed an alternative parametric procedure of combining two biomarkers to evaluate the diagnostic accuracy using bi-variate lognormality assumption. The behavior of accuracy and standard error are observed on the basis of change in correlation measure and change in parametric values. Monte Carlo simulation technique is adopted to find the approximate variance of $\hat{AUC}$. We have applied the proposed method to simulated data set as well as real life example called Prostate cancer data set. It is found that the bi-variate ROC curve lies in between the two univariate ROC curves plotted separately for each biomarker. The calculations of incomplete integral of FPR and TPR as well as standard error of AUC are calculated in R software. The proposed bi-variate lognormal ROC curve differs from other multivariate ROC methodology in the sense the proposed model is not based on any optimization strategy. We suggest the model to be employed once the data fits the bi-variate log normal distribution.