ORİJİNAL ARAŞTIRMA ORIGINAL RESEARCH

DOI: 10.5336/biostatic.2017-56589

Detecting Diagnostic Accuracy of Biomarkers Through a Bivariate Beta Distribution

İki Değişkenli Beta Dağılımı ile Biomarkırların Tanısal Doğruluğunun Tespit Edilmesi

Atanu BHATTACHARJEE,[a]
Gajendra VISHWAKARMA,[b]
Abin THOMAS[b]

[a]Centre for Cancer Epidemiology, The Advanced Centre for Treatment, Research and Education in Cancer, Tata Memorial Centre, Navi Mumbai, [b]Department of Applied Mathematics, Indian Institute of Technology, Dhanbad, INDIA

Geliş Tarihi/*Received:* 18.05.2017
Kabul Tarihi/*Accepted:* 02.08.2017

Yazışma Adresi/*Correspondence:*
Atanu BHATTACHARJEE
The Advanced Centre for Treatment, Research and Education in Cancer, Tata Memorial Centre,
Centre for Cancer Epidemiology, Navi Mumbai,
INDIA/HİNDİSTAN
atanustat@gmail.com

ABSTRACT Objective: Detection of novel biomarkers and immunotherapy targets for prostate cancer (PCa) is essential for better diagnosis and therapy. It is a challenge to establish the influence of a biomarker for disease management of prostate cancer (PCa). **Material and Methods:** In this study gene expression data of normal and cancerous human prostate tissues are used to explore the influence of biomarkers. The most frequently overexpressed two genes are considered as biomarkers. The combination of these two biomarkers by an optimal linear combination is considered to establish the cutoff value. **Results:** In this paper, we propose a method to deal with two correlated continuous biomarkers by bivariate modeling through ROC curve under Beta distribution assumption. The proposed method is applied to simulated data set and applied into prostate cancer gene biomarkers. **Conclusion:** The optimum value of bivariate Beta distribution on example data is obtained through estimates of AUC on combine accuracy of the two biomarkers together. The simulation studies have been conducted to obtain the estimates of different correlation coefficient starting from 0 to 0.9.

Keywords: ROC curves; area under the curve; sensitivity; specificity; risk prediction; Kaplan-Meier Estimator

ÖZET Amaç: Prostat kanseri (PCa) için immunoterapi hedeflerinin ve yeni biomarkırların tespit edilmesi, daha iyi teşhis ve tedavi için gereklidir. Prostat kanseri (PCa) hastalık yönetimi için biomarkırların etkisinin belirlenmesi zorlu bir iştir. **Gereç ve Yöntemler:** Bu çalışmada biomarkırların etkisini araştırmak için kanserli ve normal insan prostat dokularının gen ekspresyonları kullanılmıştır. En fazla okunan iki gen biomarkır olarak ele alınmıştır. Bu iki biomarkırın optimal linear kombinasyonu kesim değeri olarak ele alınmıştır. **Bulgular:** Bu makalede, Beta dağılımı varsayımı altında iki modelli ROC eğrisi aracılığıyla iki sürekli ve ilişkili biomarkırla ilgilenen bir yöntem önerdik. Önerilen yöntem simüle edilen veri setine ve prostat kanser geni biomarkırlarına uygulanmıştır. **Sonuç:** İki değişkenli Beta dağılımına sahip örnek veri setinin optimum değeri iki biomarkırın doğruluklarının AUC tahminleri ile birleştirilmesi ile elde edilmiştir. Simülasyon çalışmaları 0-0.9 aralığında farklı korelasyon katsayı tahminleri kullanılarak yapılmıştır.

Anahtar Kelimeler: ROC eğrileri; eğri altında kalan alan; tanımlayıcılık; duyarlılık; risk tahmini; Kaplan-Meier tahmini

A biomarker is a biologic feature that can be used to detect the presence or the progress of a disease. It is important to define the threshold limit value (TLV) of a biomarker in diagnostic medicine. The TLV is a level of the biomarker to which it is believed that a patient can be free from a disease. The TLV of a biomarker should be accurate to define the presence or absence of a disease status. The diagnostic accuracy

of a biomarker to correctly detect the disease status can be defined as sensitivity (i.e., true positive rate) and specificity (i.e., true negative rate).[1] The subject is labeled as diseased when the measurement of specific biomarker becomes greater than the TLV. The Biomarker specific TLV is required to be established for disease status detection. An important field of research serves literature with several methodologies of Bivariate and Multivariate distributions through marginal and conditional distributions.[2-5,7] Different form of Bivariate Beta distributions have been proposed in the statistics literature.[6,7,9,10,27] The application of Univariate and Bivariate Beta distributions are widely adopted in different fields. These are found suitable in population genetics, linkage analysis and drought intensity data analysis.[11-13] These are used to find out the proportions of diseased second premolars and molars in dentistry and estimation of tree diameter in forestry and for retinal image recognition measurements.[14-16] The joint density function of Bivariate Beta distribution is applied to explore the readership of two monthly magazines.[17] Recently the different class of Beta distributions are explored and has received a growing interest.[18,19] It is one of the most widely adopted distribution in experimental research.[20-22] Univariate Beta-distribution is easy to implement. It can be applied through standard Gamma-distribution.[23,24] However, work with pairs of correlated beta-distribution is difficult to conduct due to non-availability of the multivariate form of the univariate Beta distribution.[24] The marginal distribution functions and prior knowledge about a measurement of association are suitable to deal with such problems.[25] Although it is limited with near normal and weakly correlated Beta distributions.[24] The algorithm for generating data from bivariate Beta random variables has been attempted through three parameters.[26] The generation of data on correlated bivariate Beta-distributed has been explored through mixture distribution approach.[28]

The gene-biomarker expression values on prostate cancer patients are considered as diagnosis marker for illustration of correlated Beta Bivariate distribution.[27,29,30,32] The proposed method is performed through MCMC technique to discriminate the TLV of gene -biomarker expressions among healthy and diseased individual. The specific Bivariate Beta distributions are studied and finally some empirical applications with gene-biomarker expression data are presented to obtain the TLV.

MATERIAL AND METHODS

INITIAL ASSUMPTION

A Beta distribution is assumed with parameter p_1 and other known parameters α and β. The random number is assumed with k times of 'successes' that is drawn from a binomial distribution (likelihood) with parameters $p_1$and v. In next step a random draw p_2 is attempted from a Beta-distribution with parameters p_1+ k and β+v-k to obtain the paired Beta-distributed random variables. The specified correlation $v \times (v + \alpha + \beta)^{-1}$ is utilized to generate the paired Beta-distribution. The estimation of p1 and p2 were obtained through determination of$\alpha;\beta$; k and v respectively.

BIVARIATE BETA DISTRIBUTION

Suppose the gamma variates G_0, G_1, G_2 are independent having parameters $(\alpha_0, \beta_0), (\alpha_1, \beta_1)$, and (α_2, β_2)respectively. Then the joint density is obtained from

$$X = \frac{G_1}{G_1+G_2}, Y = \frac{G_2}{G_2+G_0} \quad (1)$$

TABLE 1: Estimated of C of bivariate beta for different shape and scale parameters.

$\alpha_1 = 1.0$ and $\alpha_2 = 1.0$														
β_1	β_2	C	β_1	β_2	C	β_1	β_2	C	β_1	β_2	C	β_1	β_2	C
1	1	0.20	1.5	1	0.21	2	1	0.22	2.5	1	0.23	3	1	0.23
1	1.5	0.21	1.5	1.5	0.23	2	1.5	0.24	2.5	1.5	0.25	3	1.5	0.25
1	2.0	0.22	1.5	2	0.24	2	2	0.25	2.5	2	0.26	3	2	0.26
1	2.5	0.23	1.5	2.5	0.25	2	2.5	0.26	2.5	2.5	0.26	3	2.5	0.27
1	3	0.23	1.5	3	0.25	2	3	0.26	2.5	3	0.27	3	3	0.27
$\alpha_1 = 1.0$ and $\alpha_2 = 1.5$														
β_1	β_2	C	β_1	β_2	C	β_1	β_2	C	β_1	β_2	C	β_1	β_2	C
1	1	0.21	1.5	1	0.23	2	1	0.24	2.5	1	0.25	3	1	0.25
1	1.5	0.23	1.5	1.5	0.25	2	1.5	0.26	2.5	1.5	0.26	3	1.5	0.27
1	2.0	0.24	1.5	2	0.26	2	2	0.27	2.5	2	0.27	3	2	0.28
1	2.5	0.24	1.5	2.5	0.26	2	2.5	0.27	2.5	2.5	0.28	3	2.5	0.29
1	3	0.24	1.5	3	0.26	2	3	0.28	2.5	3	0.28	3	3	0.29
$\alpha_1 = 1.0$ and $\alpha_2 = 2.0$														
β_1	β_2	C	β_1	β_2	C	β_1	β_2	C	β_1	β_2	C	β_1	β_2	C
1	1	0.22	1.5	1	0.24	2	1	0.25	2.5	1	0.26	3	1	0.26
1	1.5	0.24	1.5	1.5	0.26	2	1.5	0.27	2.5	1.5	0.27	3	1.5	0.28
1	2.5	0.24	1.5	2	0.26	2	2	0.27	2.5	2	0.28	3	2	0.29
1	2.5	0.25	1.5	2.5	0.27	2	2.5	0.28	2.5	2.5	0.29	3	2.5	0.29
1	3	0.25	1.5	3	0.27	2	3	0.28	2.5	3	0.29	3	3	0.30
$\alpha_1 = 1.0$ and $\alpha_2 = 2.5$														
β_1	β_2	C	β_1	β_2	C	β_1	β_2	C	β_1	β_2	C	β_1	β_2	C
1	1	0.23	1.5	1	0.25	2	1	0.26	2.5	1	0.26	3	1	0.27
1	1.5	0.24	1.5	1.5	0.26	2	1.5	0.27	2.5	1.5	0.28	3	1.5	0.29
1	2.0	0.25	1.5	2	0.27	2	2	0.28	2.5	2	0.29	3	2	0.29
1	2.5	0.25	1.5	2.5	0.27	2	2.5	0.28	2.5	2.5	0.29	3	2.5	0.30
1	3	0.25	1.5	3	0.27	2	3	0.29	2.5	3	0.30	3	3	0.30
$\alpha_1 = 1.0$ and $\alpha_2 = 3.0$														
β_1	β_2	C	β_1	β_2	C	β_1	β_2	C	β_1	β_2	C	β_1	β_2	C
1	1	0.23	1.5	1	0.25	2	1	0.26	2.5	1	0.27	3	1	0.27
1	1.5	0.24	1.5	1.5	0.26	2	1.5	0.28	2.5	1.5	0.28	3	1.5	0.29
1	2.0	0.25	1.5	2	0.27	2	2	0.28	2.5	2	0.29	3	2	0.30
1	2.5	0.25	1.5	2.5	0.27	2	2.5	0.29	2.5	2.5	0.30	3	2.5	0.30
1	3	0.25	1.5	3	0.27	2	3	0.29	2.5	3	0.30	3	3	0.30

and the generalized Beta distribution with density is

$$f(x,y) = \frac{1}{B(\alpha_0,\alpha_1,\alpha_2)} \frac{\lambda_1^{\alpha_1} x^{\alpha_1-1}(1-x)^{-(\alpha_1+1)} \lambda_2^{\alpha_2} x^{\alpha_2-1}(1-x)^{-(\alpha_2+1)}}{[1+\frac{\lambda_1 x}{1-x}+\frac{\lambda_2 y}{1-y}]^{\alpha_0+\alpha_1+\alpha_2}} \quad (2)$$

for$0 < x, y < 1; \alpha_i \beta_i > 0$ for i = 0,1,2, and $\lambda_i = \frac{\beta_1}{\beta_0}$for i=1,2. In (1.2) $B(\alpha_1, \dots \dots \dots, \alpha_k) = \prod \frac{\Gamma(\alpha_i)}{\Gamma(\sum \alpha_i)}$

Is the generalized beta function. When $\lambda_i = 1$, the density (2) reduces to a Bivariate Beta distribution with parameters by:

$$f(x,y) = \frac{1}{B(\alpha_0,\alpha_1,\alpha_2)} \frac{x^{\alpha_1-1}(1-x)^{\alpha_0+\alpha_2-1} y^{\alpha_2-1}(1-y)^{\alpha_0+\alpha_1-1}}{(1-xy)^{\alpha_0+\alpha_1+\alpha_2}} \quad (3)$$

The correlation coefficient between X and Y is range from [0; 1][9]. Suppose V;W, are beta random variates and

their relation with X and Y with $V = X/(1-X) = (G_1 + G_3)/(G_4 + G_5)$

and $= Y/(1-Y) = (G_2 + G_4)/(G_3 + G_5)$.

In this context, bivariate Beta distributions with five independent gamma variates$[G_1, G_2, G_3, G \text{ and} G_5]$

and common scale parameter 1 is defined by Arnold and Ng[3] and it is considered as It is defined as

$$X = \frac{G_1+G_3}{G_1+G_3+G_4+G_5} \quad (4)$$

$$Y = \frac{G_2+G_4}{G_2+G_3+G_4+G_5} \quad (5)$$

This joint density is free from closed form and can be calculated numerically. In this work, we applied the correlation structure through consideration of δ. The value of δ finalized with simulation. These random variables have distributions which are usually called Beta distribution of the second kind. The joint probability density

function (p.d.f) of(V, W, G_3, G_4, G_5) is readily verified to be of the form:

$$f(v,w,g,g,g) = \frac{(g_3+g_5)(g_4+g_5)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\alpha_3)\Gamma(\alpha_4)\Gamma(\alpha_5)}[v(g_4+g_5)-g_3]^{\alpha_1-1}$$
$$\times [w(g_3+g_5)-g_4]^{\alpha_2-1} g_3^{\alpha_3-1} g_4^{\alpha_4-1} g_5^{\alpha_5-1} \exp\{-[g_3 w + g_4 w + g_5(v+w+1)]\}\frac{g_4}{w} - g_5$$
$$< g_3 < (g_4+g_5)v, g_4 > 0, g_5 > 0, v > 0, w > 0$$

Therefore, the joint p.d.f of (V,W) is obtained as

$$f_{V,W}(v,w) = \int_0^\infty \int_0^\infty \int_{\frac{g_4}{w}-g_5}^{(g_4+g_5)v} f(v,w,g_3,g_4,g_5) dg_3 dg_4 dg_5, v, w > 0$$

And the joint p.d.f. of (X,Y) is

$$f_{X,Y}(x,y) = \frac{1}{(1-x)^2(1-y)^2} f_{V,W}\left(\frac{x}{1-x}, \frac{y}{1-y}\right), 0 < x < 1, 0 < y < 1 \quad (7)$$

CORRELATION BETWEEN BIVARIATE BETA DISTRIBUTIONS

Let the $biomarker_1$ is defined as $X_1$and $biomarker_2$ (X_2) follows Beta distribution. The distribution is defined as $Bet(\alpha_1, \beta_1)$and $t(\alpha_2, \beta_2)$.

It is defined as

$$X = \frac{G(\alpha)}{G(\alpha)+G(\beta)} \quad (8)$$

where the random variable $G(\alpha)$is distributed with standard Gamma distribution with parameters α and $1\big(Gam(\alpha;1)\big)$. Further the independent random variable $G(\beta)$ is a standard Gamma distribution having parameters $\beta$and $1\big(Gam(\alpha;1)\big)$.

It is defined as $E(X_1) = p_1 = \alpha_1/(\alpha_1+\beta_1)$and $E(X_1) = p_2 = \alpha_2/(\alpha_2+\beta_2)$.

Further, $Var(X_1) = p_1(1-p_1) = (1+\alpha_1+\beta_1)$ and $Var(X_2) = p_2(1-p_2) = (1+\alpha_2+\beta_2)$, respectively.

The correlated random variable is handled through shared random variable technique, that further extended with correlated binary variables.[31,33] Let the sum of two independent random standard Gamma variables is called as $G(\alpha_1^*) + G(\partial_1^*)$ is distributed with $Gam(\alpha_1^* + \delta_1, 1)$.[34,35]

Now,

$$X_1 = \frac{G(\alpha_1^*)+G(\delta_1)}{G(\alpha_1^*)+G(\delta_1)+G(\beta_1^*)+G(\delta_2)} \quad (9)$$

$$X_2 = \frac{G(\alpha_2^*)+G(\delta_1)}{G(\alpha_2^*)+G(\delta_1)+G(\beta_2^*)+G(\delta_2)} \quad (10)$$

The terms δ_1 and δ_2 are linked with $\alpha_1, \alpha_2, \beta_1$and β_2 as $\alpha_1^* + \delta_1 = \alpha_1, \alpha_2^* + \delta_1 = \alpha_2, \beta_1^* + \delta_2 = \beta_1$,and $\beta_2^* + \delta_2 = \beta_2$, respectively for the marginal distributions of X_1 and X_2.[35] The covariance (ρ_{xx}) between X_1 and X_2 is approximated with first-ordered Taylor series through

$$\rho = E\left(\frac{dX_1}{G(\delta_1)} \times \frac{dX_2}{G(\delta_1)}\right) \times var(G(\delta_1)) + E(\frac{dX_1}{G(\delta_2)} \times \frac{dX_2}{G(\delta_2)}) \times var(G(\delta_2)) \quad (11)$$

where, E denotes the expected value operator. The simplified form of this expression is

$$\rho_{xx} = \frac{\alpha_1 \times \alpha_2 \times \delta_2 + (1+\beta_1) \times (1+\beta_2) \times \delta_1}{(\alpha_1+\beta_1) \times (\alpha_2+\beta_2) \times (1+\alpha_1+\beta_1) \times (1+\alpha_2+\beta_2)} \quad (12)$$

The terms $\hat{\partial}_1$ and $\hat{\partial}_2$ are estimated through

$$\hat{\partial}_1 = \rho_{xx} \times (1 + \alpha_1 + \alpha_2) \times C \quad (13)$$

$$\hat{\partial}_2 = \rho_{xx} \times (1 + \beta_1 + \beta_2\ ) \times C \quad (14)$$

where

$$C = \frac{\sqrt{\alpha_1 \times \alpha_2 \times \beta_1 \times \beta_2 \times (1+\alpha_1+\beta_1) \times (1+\alpha_2+\beta_2)}}{(1+\alpha_2) \times (1+\beta_1) \times (1+\beta_2) + \alpha_1(1+\beta_1+\beta_2+\beta_1\beta_2+\alpha_2(1+\beta_1+\beta_2))} \quad (15)$$

DATA METHODOLOGY

The proposed extensions of the bivariate Beta distribution model was applied to real data set, publicly available in the GEO(Gene Expression Omnibus) database. The GEO accession number of the data set is GDS4824. A total of 21 samples (13 prostate cancer patients and 8 normal individuals) are considered in this study. The study was about gene expression of malignant and benign stage of prostate tissues. It provides insight about gene expression signature for prostate cancer. The objective of the primary study was the Identification of novel biomarkers and immunotherapy targets for prostate cancer (PCa) is crucial to better diagnosis and therapy. The library ("limma") in R (open source software) is used to explore and tabled the most significant genes ID.[36] Initially, all gene expression changes among cancer and normal individuals are observed and given in Figure 1 through Heatmap representation. It has been observed through colors that few genes were expressed significantly different between malignant and benign tissues. Thereafter significant level cut off the value of *p=0.05* is considered to be labeled the most differently expressed genes and that are given in Figure 2. In next step, a significant level was decided with $p=0.001$ is selected to detect the genes those expressions were highly different between malignant and benign tissues. It has been observed only genes (Figure 3) were differently expressed malignant and benign tissues. Only Two genes from there i.e (gene ID with 209426 and 207147), were considered in this study for computation simplicity and AUC estimation.

BAYESIAN ANALYSIS

Suppose the pair of independent random variables areK_{1i} and K_{2i}.It is assumed that

$K_{1i} \sim Binomial\ (n_{1i}, p_1)andK_{2i} \sim Binomial\ (n_{2i}, p_2), i = 1,2, \ldots . N$

The parameters to be estimated are p_1 and p_2 respectively. It is assumed that p_1 and $p_2$are correlated in prior beliefs. Let $K_1 = \sum_{i=1}^{N} K_{1i}$ and $K_2 = \sum_{i=1}^{N} K_{2i}, n_1 = \sum_{i=1}^{N} n_{1i}$ and $n_2 = \sum_{i=1}^{N} n_{2i}$. The prior distribution of the parameters p_1 and p_2 are from (4) with $(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5)$ and

$$\Pr(K_1 = k_1, K_2 = k_2 | p_1, p_2) = \binom{n_1}{k_1} {p_1}^{k_1} (1-p_1)^{n_1-k_1} \binom{n_2}{k_2} {p_2}^{k_2} (1-p_2)^{n_2-k_2} \quad (16)$$

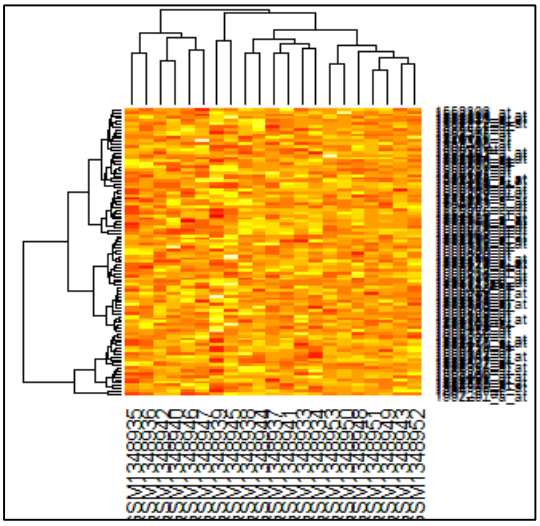

FIGURE 1: Heat map showing identification of the transcription factor as a potential biomarker and immunotherapy target in prostate cancer patietns in comparison normal individuals.

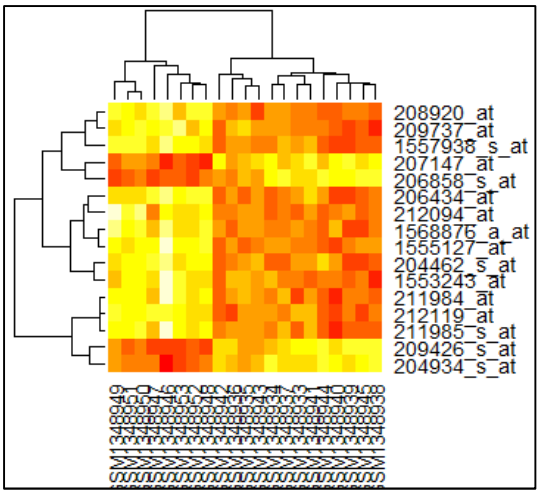

FIGURE 2: Heat map showing selected genes(with p value 0.05) as a potential biomarker and immunotherapy target in prostate cancer patients in comparison normal individuals.

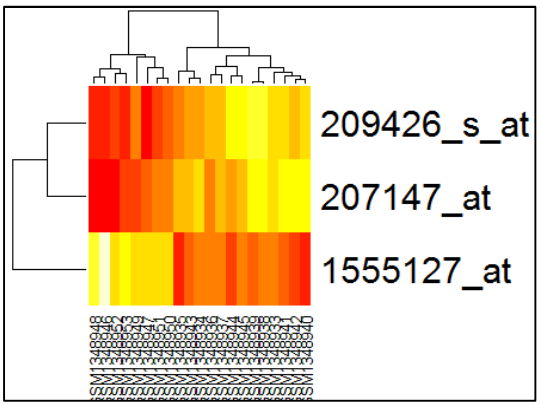

FIGURE 3: Heat map showing selected genes(with p value 0.05) as a potential biomarker and immunotherapy target in prostate cancer patietns in comparison normal individuals.

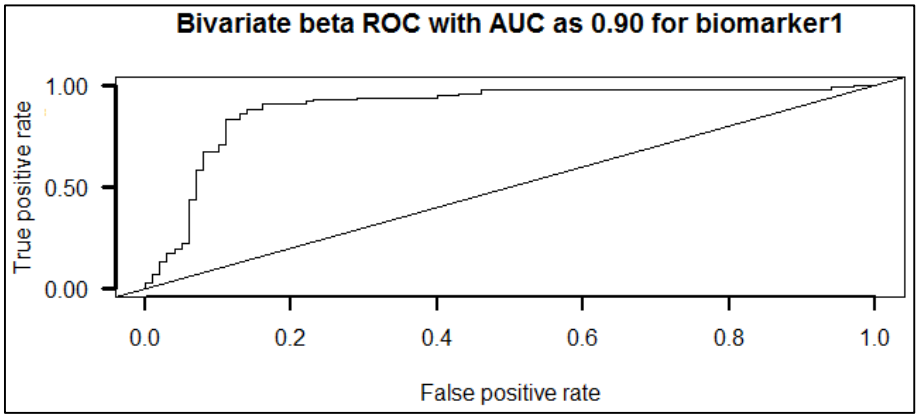

FIGURE 4:-BetaROC curve for biomarker 1 with AUC as 0.90.

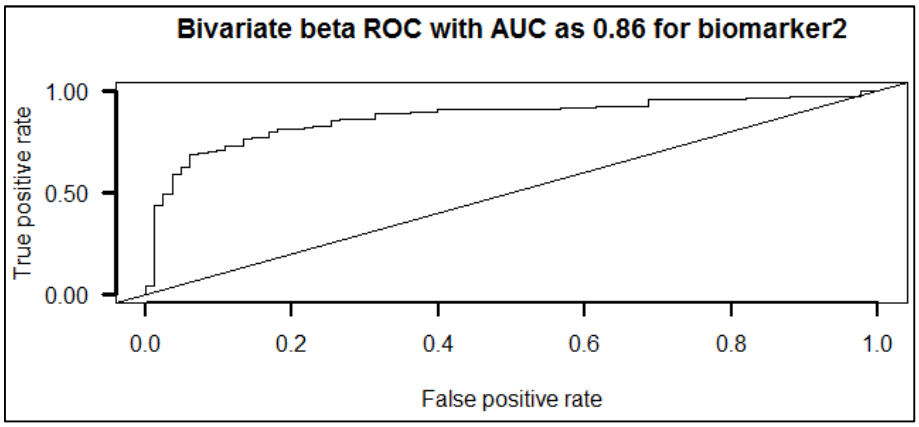

FIGURE 5:- BetaROC curve for biomarker 1 with AUC as 0.86.

where $k_1 = 0,1,2,\ldots\ldots n_1$,$k_2 = 0,1,2,\ldots\ldots n_2$

Further,

$$f(p_1,p_2|K_1 = k_1, K_2 = k_2) = \frac{Pr(K_1=k_1,K_2=k_2|p_1,p_2\,)f_{P_1,P_2}(p_1,p_2)}{\int_0^t\int_0^t Pr(K_1=k_1,K_2=k_2|p_1,p_2)\,f_{P_1,P_2}(p_1,p_2)dp_1dp_2} \quad (17)$$

where $0 < p_1 < 1,0 < p_2 < 1$

The expected value of posteriors of p_1 and p_2 are obtained through

$$E(p_1) = \int_0^t\int_0^t p_1 f(p_1,p_2|K_1 = k_1, K_2 = k_2)dp_1dp_2 \quad (18)$$

and

$$E(p_2) = \int_0^t\int_0^t p_2 f(p_1,p_2|K_1 = k_1, K_2 = k_2)dp_1dp_2 \quad (19)$$

The random variables are obtained from $f_{P_1,P_2}(p_1,p_2)$ for specified prior values of $(\alpha_1,\alpha_2,\alpha_3,\alpha_4,\alpha_5), i = 1,2,\ldots\ldots m.$ and thereafter it approximated with integrals by

$$A_1 = \int_0^1\int_0^1 \Pr(K_1 = k_2, K_2 = k_2|p_1,p_2)\,dp_1dp_2 \quad (20)$$

$$A_1 \approx \frac{C}{m}\sum_{j=1}^m {p_1}_j{}^{k_1}(1-{p_1}_j)^{n_1-k_1}{p_2}_i{}^{k_2}(1-{p_2}_j)^{n_2-k_2} \quad (21)$$

$$A_2 = \int_0^1\int_0^1 p_1 f(p_1,p_2|K_1 = k_1, K_2 = k_2)dp_1dp_2 \quad (22)$$

$$A_2 \approx \frac{C}{A_1 m}\sum_{j=1}^{m} {p_{1j}}^{k_1}(1-p_{1j})^{n_1-k_1}{p_{2i}}^{k_2+1}(1-p_{2j})^{n_2-k_2} \quad (23)$$

where $C = \binom{n_1}{k_1}\binom{n_2}{k_2}$

The $100(1-\gamma)\%$ highest posterior density (HPD) interval for the parameter $p_1, (L_1, U_1)$ is obtained for the parameters through

$$\int_0^1 \int_0^{U_1} f(p_1, p_2 | K_1 = k_1, K_1 = k_2) dp_1 dp_2 = 1 - \gamma/2 \quad (24)$$

and

$$\int_0^1 \int_0^{L_1} f(p_1, p_2 | K_1 = k_1, K_1 = k_2) dp_1 dp_2 = \gamma/2 \quad (25)$$

The Monte Carlo integration is used with interval (0,1) to approximate the values of U_1 and L_1.

The estimation for p_2 is obtained through $100(1-\gamma)\%$ highest posterior density (HPD) interval. Similarly, for p_1 the HPD is obtained through $100(1-\gamma)\%$. Initial data of (p_1, p_2) are obtained from $f_{P_1,P_2}(p_1, p_2)$ with parameter$(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5)$.

In next step it is used to generate values of $K_1, K_2)$,say (k_1, k_2). The Monte Carlo simulation is used to evaluate the parameters 20,000 simulations are used for each setting with sample size n=50 and 100. The simulated biases and mean squared errors (MSEs) of the estimators are presented in Table 2 and Table 3 for two representative sets of parameter $(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5)$ =(1,1.5,2,2.5,3) and (1,1.5,2,2.5,3), respectively. It is to be noted that the biases and MSEs decrease when the sample size increases.

ROC ANALYSIS

Sometimes two biomarkers may be highly associated with the presence of disease and it is indispensable to consider both the biomarkers for detecting the disease status.[38] Let (X_1, X_2) be two sets of related biomarkers taken from a healthy group and let (Y_1, Y_2) be two sets of biomarkers value taken from disease subjects. A particular subject is identified as disease when the values of Y_1 and Y_2 are large enough.[38] The CDF for healthy and disease population are defined by $F_x(t_1, t_2) = P(X_1 \le t_1, X_2 \le t_2)$ and $F_y(t_1, t_2) = P(Y_1 \le t_1, Y_2 \le t_2)$ respectively. The FPR and TPR in the bivariate 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 continous biomarker values from the healthy and disease groups, respectively.

Further, it is assumed that $F_x(t_1, t_2) \sim Be(\alpha_1, \beta_1)$ and $F_y(t_1, t_2) \sim Be(\alpha_2, \beta_2)$. The estimation procedure for parameters above is obtained from mean and variance of Beta distribution for each sample. The mean and variance observed separately disease and not disease individuals. The moment approach is easy to use for the iterative solution for the maximum likelihood estimates. It is assumed for the disease individual, the mean is standard deviations is σ_x^2. Similarly, from the sample data in normal individual, the mean $\hat{p}_2$ and standard deviations is σ_y^2 respectively, the estimates of the parameters is

$$\hat{\alpha}_1 = \bar{p}_1\{\bar{p}_1(1-\bar{p}_1)\sigma_x^2 - 1\}, \hat{\alpha}_2 = \bar{p}_2\{\bar{p}_2(1-\bar{p}_2)\sigma_y^2 - 1\},$$

$$\widehat{\beta_1} = (1-\overline{p_1})\{\overline{p_1}(1-\overline{p_1})\sigma_x^2 - 1\}, \widehat{\beta_2} = (1-\overline{p_2})\{\overline{p_2}(1-\overline{p_2})\sigma_y^2 - 1\}$$

The AUC of bivariate beta ROC curve takes the following form:

$$\text{AUC} = \text{P}(\text{Y}_1 > \text{X}_1, \text{Y}_2 > \text{X}_2) \int\int \text{I}(\text{y}_1 > \text{x}_1, \text{y}_2 > \text{x}_2) \text{dF}_\text{X}(\text{t}_1, \text{t}_2) \text{dF}_\text{Y}(\text{t}_1, \text{t}_2) \quad (26)$$

TABLE 2: Estimated of C of bivariate beta for different shape and scale parameters.

$\alpha_1 = 1.5$ and $\alpha_2 = 1.0$														
β_1	β_2	C	β_1	β_2	C	β_1	β_2	C	β_1	β_2	C	β_1	β_2	C
1	1	0.21	1.5	1	0.23	2	1	0.24	2.5	1	0.24	3	1	0.24
1	1.5	0.23	1.5	1.5	0.25	2	1.5	0.26	2.5	1.5	0.26	3	1.5	0.26
1	2.0	0.24	1.5	2	0.26	2	2	0.27	2.5	2	0.27	3	2	0.28
1	2.5	0.25	1.5	2.5	0.26	2	2.5	0.27	2.5	2.5	0.28	3	2.5	0.28
1	3	0.25	1.5	3	0.27	2	3	0.28	2.5	3	0.29	3	3	0.29
$\alpha_1 = 1.5$ and $\alpha_2 = 1.5$														
β_1	β_2	C	β_1	β_2	C	β_1	β_2	C	β_1	β_2	C	β_1	β_2	C
1	1	0.23	1.5	1	0.25	2	1	0.26	2.5	1	0.26	3	1	0.26
1	1.5	0.25	1.5	1.5	0.26	2	1.5	0.27	2.5	1.5	0.28	3	1.5	0.28
1	2.0	0.26	1.5	2	0.27	2	2	0.29	2.5	2	0.29	3	2	0.30
1	2.5	0.26	1.5	2.5	0.28	2	2.5	0.29	2.5	2.5	0.30	3	2.5	0.31
1	3	0.26	1.5	3	0.28	2	3	0.30	2.5	3	0.31	3	3	0.31
$\alpha_1 = 1.5$ and $\alpha_2 = 2.0$														
β_1	β_2	C	β_1	β_2	C	β_1	β_2	C	β_1	β_2	C	β_1	β_2	C
1	1	0.22	1.5	1	0.26	2	1	0.27	2.5	1	0.27	3	1	0.28
1	1.5	0.26	1.5	1.5	0.27	2	1.5	0.29	2.5	1.5	0.29	3	1.5	0.30
1	2.5	0.26	1.5	2	0.28	2	2	0.30	2.5	2	0.30	3	2	0.31
1	2.5	0.27	1.5	2.5	0.29	2	2.5	0.30	2.5	2.5	0.31	3	2.5	0.32
1	3	0.27	1.5	3	0.29	2	3	0.31	2.5	3	0.31	3	3	0.32
$\alpha_1 = 1.5$ and $\alpha_2 = 2.5$														
β_1	β_2	C	β_1	β_2	C	β_1	β_2	C	β_1	β_2	C	β_1	β_2	C
1	1	0.25	1.5	1	0.26	2	1	0.26	2.5	1	0.26	3	1	0.27
1	1.5	0.26	1.5	1.5	0.28	2	1.5	0.27	2.5	1.5	0.28	3	1.5	0.29
1	2.0	0.27	1.5	2	0.28	2	2	0.28	2.5	2	0.29	3	2	0.29
1	2.5	0.27	1.5	2.5	0.29	2	2.5	0.28	2.5	2.5	0.29	3	2.5	0.30
1	3	0.27	1.5	3	0.30	2	3	0.29	2.5	3	0.30	3	3	0.30
$\alpha_1 = 1.5$ and $\alpha_2 = 3.0$														
β_1	β_2	C	β_1	β_2	C	β_1	β_2	C	β_1	β_2	C	β_1	β_2	C
1	1	0.25	1.5	1	0.27	2	1	0.28	2.5	1	0.27	3	1	0.27
1	1.5	0.26	1.5	1.5	0.28	2	1.5	0.31	2.5	1.5	0.28	3	1.5	0.29
1	2.0	0.27	1.5	2	0.29	2	2	0.31	2.5	2	0.29	3	2	0.30
1	2.5	0.27	1.5	2.5	0.30	2	2.5	0.31	2.5	2.5	0.30	3	2.5	0.30
1	3	0.27	1.5	3	0.30	2	3	0.31	2.5	3	0.30	3	3	0.30

The explicit form of FPR and TPR in case of bivariate setting is analytically complicated. The FPR from a bivariate biomarker at the thresholds t_1 and t_2 are defined for TLVs for respective biomarkers. The FPR and TPR is presented through Beta distribution.

Let the density function X is defined as

$$f_x = \frac{1}{B(\alpha_1,\beta_1)} x^{\alpha_1-1}(1-x)^{\beta_1-1}, \alpha_1 \geq 1, \beta_1 = 1 \quad (27)$$

and

$$f_y = \frac{1}{B(\alpha_2,\beta_2)} y^{\alpha_2-1}(1-y)^{\beta_2-1}, \alpha_2 = 1, \beta \geq 1 \quad (28)$$

The term B(.) provides a normalization constant that equals to $\frac{1}{\alpha_1}$ and $\frac{1}{\beta_1}$ for equations with the interval [0,1].[27] Now subscripts of (β_1, α_1) and (β_2, α_2) are presented as (β, α) respectively. Integrating over the equation from 0 to t_1 gives the following

$$FPR = x^{\alpha_1} \quad (29)$$

$$TPR = 1 - (1-x)^{\beta_1} \quad (30)$$

Similarly, the equation (28) provides

TABLE 3. Estimated of C of bivariate beta for different shape and scale parameters.

$\alpha_1 = 2.0$ and $\alpha_2 = 1.0$														
β_1	β_2	C	β_1	β_2	C	β_1	β_2	C	β_1	β_2	C	β_1	β_2	C
1	1	0.22	1.5	1	0.24	2	1	0.24	2.5	1	0.25	3	1	0.25
1	1.5	0.24	1.5	1.5	0.26	2	1.5	0.26	2.5	1.5	0.27	3	1.5	0.27
1	2.0	0.25	1.5	2	0.27	2	2	0.27	2.5	2	0.28	3	2	0.28
1	2.5	0.26	1.5	2.5	0.27	2	2.5	0.28	2.5	2.5	0.29	3	2.5	0.29
1	3	0.26	1.5	3	0.28	2	3	0.29	2.5	3	0.29	3	3	0.30
$\alpha_1 = 2.0$ and $\alpha_2 = 1.5$														
β_1	β_2	C	β_1	β_2	C	β_1	β_2	C	β_1	β_2	C	β_1	β_2	C
1	1	0.24	1.5	1	0.26	2	1	0.26	2.5	1	0.27	3	1	0.27
1	1.5	0.26	1.5	1.5	0.27	2	1.5	0.28	2.5	1.5	0.29	3	1.5	0.29
1	2.0	0.27	1.5	2	0.29	2	2	0.30	2.5	2	0.30	3	2	0.31
1	2.5	0.27	1.5	2.5	0.29	2	2.5	0.30	2.5	2.5	0.31	3	2.5	0.31
1	3	0.28	1.5	3	0.30	2	3	0.31	2.5	3	0.32	3	3	0.32
$\alpha_1 = 2.0$ and $\alpha_2 = 2.0$														
β_1	β_2	C	β_1	β_2	C	β_1	β_2	C	β_1	β_2	C	β_1	β_2	C
1	1	0.24	1.5	1	0.26	2	1	0.26	2.5	1	0.27	3	1	0.27
1	1.5	0.26	1.5	1.5	0.27	2	1.5	0.28	2.5	1.5	0.29	3	1.5	0.29
1	2.5	0.27	1.5	2	0.28	2	2	0.30	2.5	2	0.30	3	2	0.31
1	2.5	0.27	1.5	2.5	0.29	2	2.5	0.30	2.5	2.5	0.31	3	2.5	0.31
1	3	0.28	1.5	3	0.30	2	3	0.31	2.5	3	0.32	3	3	0.32
$\alpha_1 = 2.0$ and $\alpha_2 = 2.5$														
β_1	β_2	C	β_1	β_2	C	β_1	β_2	C	β_1	β_2	C	β_1	β_2	C
1	1	0.25	1.5	1	0.27	2	1	0.27	2.5	1	0.28	3	1	0.28
1	1.5	0.27	1.5	1.5	0.29	2	1.5	0.30	2.5	1.5	0.30	3	1.5	0.31
1	2.0	0.27	1.5	2	0.30	2	2	0.31	2.5	2	0.31	3	2	0.32
1	2.5	0.28	1.5	2.5	0.30	2	2.5	0.31	2.5	2.5	0.32	3	2.5	0.33
1	3	0.28	1.5	3	0.31	2	3	0.32	2.5	3	0.33	3	3	0.33
$\alpha_1 = 2.0$ and $\alpha_2 = 3.0$														
β_1	β_2	C	β_1	β_2	C	β_1	β_2	C	β_1	β_2	C	β_1	β_2	C
1	1	0.26	1.5	1	0.27	2	1	0.29	2.5	1	0.29	3	1	0.30
1	1.5	0.28	1.5	1.5	0.30	2	1.5	0.31	2.5	1.5	0.32	3	1.5	0.32
1	2.0	0.28	1.5	2	0.31	2	2	0.32	2.5	2	0.33	3	2	0.33
1	2.5	0.29	1.5	2.5	0.31	2	2.5	0.33	2.5	2.5	0.33	3	2.5	0.34
1	3	0.29	1.5	3	0.31	2	3	0.33	2.5	3	0.34	3	3	0.35

$$FPR = x^{\alpha_2} \tag{31}$$

$$TPR = 1 - (1 - x)^{\beta_2} \tag{32}$$

The relation between TFR and FPR has been established as $TPR = 1 - (1 - FPR^{\frac{1}{\alpha}})^{\beta}, \alpha \geq 1, \beta \geq 1$ (33)
The area under a ROC is considered from FPR=0 to FPR=θ is

where $B(\theta^{\frac{1}{\alpha}}, \alpha, \beta + 1)$ is the incomplete beta function. Further AUC is extended as

$$AUC = \int_0^t 1 - (1 - FPR^{\frac{1}{\alpha}})^{\beta} dFPR = \theta - \alpha B(\theta^{\frac{1}{\alpha}}, \alpha, \beta + 1) \tag{33}$$

SIMULATION STUDY

Initially we explore with different values of α_i and β_i to obtained the possible estimates of $\hat{C}$ .The simulated values of $\hat{C}$ with different combinations of $\alpha_i$and $\beta_i$are presented in details in Table1-5. It is to be noted that the $\hat{C}$ values are quite consistent over different choice of$\alpha_i$and β_i respectively. Further, estimated values of $\hat{C}$.is used to get estimates about $\hat{\rho}$, $\hat{\delta}_1$, and $\hat{\delta}_2$.It shows that the estimates of $\hat{\delta}_1$, and $\hat{\delta}_2$ are quite less and almost near to zero as desired in theoretical section.There after the random selected values of α_i and β_i (from the frame of i=1,1.5,2,2.5,3) are considered to generate the $\widehat{AUC}$ for gene

TABLE 4. Estimated of C of bivariate beta for different shape and scale parameters.

$\alpha_1 = 2.5$ and $\alpha_2 = 1.0$														
β_1	β_2	C	β_1	β_2	C	β_1	β_2	C	β_1	β_2	C	β_1	β_2	C
1	1	0.23	1.5	1	0.24	2	1	0.25	2.5	1	0.25	3	1	0.25
1	1.5	0.25	1.5	1.5	0.26	2	1.5	0.26	2.5	1.5	0.27	3	1.5	0.27
1	2.0	0.26	1.5	2	0.27	2	2	0.27	2.5	2	0.28	3	2	0.29
1	2.5	0.26	1.5	2.5	0.28	2	2.5	0.28	2.5	2.5	0.29	3	2.5	0.30
1	3	0.27	1.5	3	0.29	2	3	0.29	2.5	3	0.30	3	3	0.30
$\alpha_1 = 2.5$ and $\alpha_2 = 1.5$														
β_1	β_2	C	β_1	β_2	C	β_1	β_2	C	β_1	β_2	C	β_1	β_2	C
1	1	0.25	1.5	1	0.26	2	1	0.27	2.5	1	0.27	3	1	0.27
1	1.5	0.26	1.5	1.5	0.28	2	1.5	0.29	2.5	1.5	0.29	3	1.5	0.30
1	2.0	0.27	1.5	2	0.29	2	2	0.30	2.5	2	0.31	3	2	0.31
1	2.5	0.28	1.5	2.5	0.30	2	2.5	0.31	2.5	2.5	0.32	3	2.5	0.32
1	3	0.28	1.5	3	0.31	2	3	0.32	2.5	3	0.32	3	3	0.33
$\alpha_1 = 2.5$ and $\alpha_2 = 2.0$														
β_1	β_2	C	β_1	β_2	C	β_1	β_2	C	β_1	β_2	C	β_1	β_2	C
1	1	0.26	1.5	1	0.27	2	1	0.28	2.5	1	0.28	3	1	0.29
1	1.5	0.27	1.5	1.5	0.29	2	1.5	0.30	2.5	1.5	0.31	3	1.5	0.31
1	2.5	0.28	1.5	2	0.30	2	2	0.31	2.5	2	0.32	3	2	0.33
1	2.5	0.29	1.5	2.5	0.31	2	2.5	0.32	2.5	2.5	0.33	3	2.5	0.33
1	3	0.29	1.5	3	0.31	2	3	0.33	2.5	3	0.34	3	3	0.34
$\alpha_1 = 2.5$ and $\alpha_2 = 2.5$														
β_1	β_2	C	β_1	β_2	C	β_1	β_2	C	β_1	β_2	C	β_1	β_2	C
1	1	0.26	1.5	1	0.28	2	1	0.29	2.5	1	0.29	3	1	0.30
1	1.5	0.28	1.5	1.5	0.30	2	1.5	0.31	2.5	1.5	0.32	3	1.5	0.32
1	2.0	0.29	1.5	2	0.31	2	2	0.32	2.5	2	0.33	3	2	0.33
1	2.5	0.29	1.5	2.5	0.32	2	2.5	0.33	2.5	2.5	0.34	3	2.5	0.34
1	3	0.30	1.5	3	0.32	2	3	0.33	2.5	3	0.34	3	3	0.35
$\alpha_1 = 2.5$ and $\alpha_2 = 3.0$														
β_1	β_2	C	β_1	β_2	C	β_1	β_2	C	β_1	β_2	C	β_1	β_2	C
1	1	0.27	1.5	1	0.29	2	1	0.29	2.5	1	0.29	3	1	0.30
1	1.5	0.28	1.5	1.5	0.31	2	1.5	0.31	2.5	1.5	0.32	3	1.5	0.33
1	2.0	0.29	1.5	2	0.31	2	2	0.31	2.5	2	0.33	3	2	0.34
1	2.5	0.30	1.5	2.5	0.32	2	2.5	0.32	2.5	2.5	0.33	3	2.5	0.35
1	3	0.30	1.5	3	0.32	2	3	0.32	2.5	3	0.34	3	3	0.35

biomarkers. The estimated value of $\widehat{AUC}$ is observed with 95% credible intervals are presented in Table 9. Based on simulated data the ROC curves are generated (Figure 1, 2).

RESULT

A total of 20,000 separate iterations is performed to obtain the estimates of $\widehat{AUC}$. The SD,95% HPD and MC error for AUC of bivariate beta ROC curve are presented in Table 9.From Table 1-8, we found that theρ is reversely proportional between biomarkers. It may be the due present correlation between biomarker is negligible or $X$and $Y$are independent. The program to generate AUC has been performed in OpenBUGS. The convergence of parameters has been observed with trace plots obtained from OpenBUGS outputs. For example Table 9 shows that for $\alpha_1 = 1, \alpha_2 = 2, \beta_1 = 2, \beta_2 = 2$, the posterior mean estimates of AUC is 0.83 with credible interval (0.81,0.86).The SD is observed with 0.03. From Table 9, we observe that as the estimated parametric values deviate more among each other especially AUC decreases with incline value of β_1.

TABLE 5. Estimated of C of bivariate beta for different shape and scale parameters.

$\alpha_1 = 3.0$ and $\alpha_2 = 1.0$														
β_1	β_2	C	β_1	β_2	C	β_1	β_2	C	β_1	β_2	C	β_1	β_2	C
1	1	0.23	1.5	1	0.24	2	1	0.25	2.5	1	0.25	3	1	0.25
1	1.5	0.25	1.5	1.5	0.26	2	1.5	0.27	2.5	1.5	0.27	3	1.5	0.27
1	2.0	0.26	1.5	2	0.28	2	2	0.28	2.5	2	0.29	3	2	0.29
1	2.5	0.27	1.5	2.5	0.28	2	2.5	0.29	2.5	2.5	0.30	3	2.5	0.30
1	3	0.27	1.5	3	0.29	2	3	0.30	2.5	3	0.30	3	3	0.30
$\alpha_1 = 3.0$ and $\alpha_2 = 1.5$														
β_1	β_2	C	β_1	β_2	C	β_1	β_2	C	β_1	β_2	C	β_1	β_2	C
1	1	0.25	1.5	1	0.26	2	1	0.27	2.5	1	0.27	3	1	0.27
1	1.5	0.27	1.5	1.5	0.28	2	1.5	0.29	2.5	1.5	0.29	3	1.5	0.30
1	2.0	0.28	1.5	2	0.29	2	2	0.30	2.5	2	0.31	3	2	0.31
1	2.5	0.29	1.5	2.5	0.30	2	2.5	0.31	2.5	2.5	0.32	3	2.5	0.32
1	3	0.29	1.5	3	0.31	2	3	0.32	2.5	3	0.32	3	3	0.33
$\alpha_1 = 3.0$ and $\alpha_2 = 2.0$														
β_1	β_2	C	β_1	β_2	C	β_1	β_2	C	β_1	β_2	C	β_1	β_2	C
1	1	0.29	1.5	1	0.28	2	1	0.28	2.5	1	0.29	3	1	0.29
1	1.5	0.28	1.5	1.5	0.30	2	1.5	0.31	2.5	1.5	0.31	3	1.5	0.31
1	2.5	0.29	1.5	2	0.31	2	2	0.32	2.5	2	0.33	3	2	0.33
1	2.5	0.29	1.5	2.5	0.32	2	2.5	0.33	2.5	2.5	0.33	3	2.5	0.34
1	3	0.30	1.5	3	0.32	2	3	0.33	2.5	3	0.34	3	3	0.35
$\alpha_1 = 3.0$ and $\alpha_2 = 2.5$														
β_1	β_2	C	β_1	β_2	C	β_1	β_2	C	β_1	β_2	C	β_1	β_2	C
1	1	0.27	1.5	1	0.28	2	1	0.29	2.5	1	0.30	3	1	0.30
1	1.5	0.29	1.5	1.5	0.31	2	1.5	0.31	2.5	1.5	0.32	3	1.5	0.32
1	2.0	0.29	1.5	2	0.32	2	2	0.33	2.5	2	0.33	3	2	0.34
1	2.5	0.30	1.5	2.5	0.32	2	2.5	0.34	2.5	2.5	0.34	3	2.5	0.35
1	3	0.30	1.5	3	0.33	2	3	0.34	2.5	3	0.35	3	3	0.35
$\alpha_1 = 3.0$ and $\alpha_2 = 3.0$														
β_1	β_2	C	β_1	β_2	C	β_1	β_2	C	β_1	β_2	C	β_1	β_2	C
1	1	0.27	1.5	1	0.29	2	1	0.30	2.5	1	0.30	3	1	0.30
1	1.5	0.29	1.5	1.5	0.31	2	1.5	0.32	2.5	1.5	0.33	3	1.5	0.33
1	2.0	0.30	1.5	2	0.32	2	2	0.33	2.5	2	0.34	3	2	0.35
1	2.5	0.30	1.5	2.5	0.33	2	2.5	0.34	2.5	2.5	0.35	3	2.5	0.35
1	3	0.30	1.5	3	0.33	2	3	0.35	2.5	3	0.35	3	3	0.36

DISCUSSION

It is a tedious job to handle with multiple diagnostic biomarkers to take the decision about diagnosis and management of a disease. The challenge is to provide the comprehensive conclusion about disease status and stage to the treated patients in the presence of multiple biomarkers through TLV. In this paper, we have proposed the method of diagnostic decision about two correlated randomly selected biomarkers. The parametric distribution is adopted to calculate the diagnostic accuracy of bivariate beta distributed biomarkers. The method is proposed for combining the biomarkers to predict the accuracy. We emphasized the application of Beta distribution over normal distribution because the range of normal distribution is $(-\infty, \infty)$.Bivariate Beta distributions are widely adopted in different applications. It is most suitable to be assumed as a prior distribution for the correlated binomial variable in Bayesian setup.[39] The prior information of two correlated random variables applied in clinical trials.[40,41]

CONCLUSION

The above-mentioned applications are not appropriate for positively correlated random variables. However in real life, it may not be possible to get only positively correlated random variables. In a

TABLE 6: Estimated of C of bivariate beta for different shape and scale parameters.

$\alpha_1 = 1.0$ and $\alpha_2 = 1.0$														
β_1	β_2	$\hat{\rho}$	β_1	β_2	$\hat{\rho}$	β_1	β_2	$\hat{\rho}$	β_1	β_2	$\hat{\rho}$	β_1	β_2	$\hat{\rho}$
1	1	0.03	1.5	1	0.02	2	1	0.02	2.5	1	0.02	3	1	0.01
1	1.5	0.02	1.5	1.5	0.02	2	1.5	0.02	2.5	1.5	0.02	3	1.5	0.01
1	2.0	0.02	1.5	2	0.02	2	2	0.02	2.5	2	0.01	3	2	0.01
1	2.5	0.02	1.5	2.5	0.02	2	2.5	0.01	2.5	2.5	0.01	3	2.5	0.01
1	3	0.01	1.5	3	0.01	2	3	0.01	2.5	3	0.01	3	3	0.01
$\alpha_1 = 1.0$ and $\alpha_2 = 1.5$														
β_1	β_2	$\hat{\rho}$	β_1	β_2	$\hat{\rho}$	β_1	β_2	$\hat{\rho}$	β_1	β_2	$\hat{\rho}$	β_1	β_2	$\hat{\rho}$
1	1	0.02	1.5	1	0.02	2	1	0.01	2.5	1	0.01	3	1	0.01
1	1.5	0.02	1.5	1.5	0.01	2	1.5	0.01	2.5	1.5	0.01	3	1.5	0.01
1	2.0	0.01	1.5	2	0.01	2	2	0.01	2.5	2	0.01	3	2	0.01
1	2.5	0.01	1.5	2.5	0.01	2	2.5	0.01	2.5	2.5	0.01	3	2.5	0.01
1	3	0.01	1.5	3	0.01	2	3	0.01	2.5	3	0.01	3	3	0.01
$\alpha_1 = 1.0$ and $\alpha_2 = 2.0$														
β_1	β_2	$\hat{\rho}$	β_1	β_2	$\hat{\rho}$	β_1	β_2	$\hat{\rho}$	β_1	β_2	$\hat{\rho}$	β_1	β_2	$\hat{\rho}$
1	1	0.01	1.5	1	0.01	2	1	0.02	2.5	1	0.01	3	1	0.01
1	1.5	0.01	1.5	1.5	0.01	2	1.5	0.02	2.5	1.5	0.01	3	1.5	0.01
1	2.0	0.01	1.5	2	0.01	2	2	0.02	2.5	2	0.01	3	2	0.01
1	2.5	0.01	1.5	2.5	0.01	2	2.5	0.01	2.5	2.5	0.01	3	2.5	0.01
1	3	0.01	1.5	3	0.01	2	3	0.01	2.5	3	0.01	3	3	0.01
$\alpha_1 = 1.0$ and $\alpha_2 = 2.5$														
β_1	β_2	$\hat{\rho}$	β_1	β_2	$\hat{\rho}$	β_1	β_2	$\hat{\rho}$	β_1	β_2	$\hat{\rho}$	β_1	β_2	$\hat{\rho}$
1	1	0.01	1.5	1	0.01	2	1	0.02	2.5	1	0.01	3	1	0.01
1	1.5	0.01	1.5	1.5	0.01	2	1.5	0.02	2.5	1.5	0.01	3	1.5	0.01
1	2.0	0.01	1.5	2	0.01	2	2	0.02	2.5	2	0.01	3	2	0.01
1	2.5	0.01	1.5	2.5	0.01	2	2.5	0.01	2.5	2.5	0.01	3	2.5	0.01
1	3	0.01	1.5	3	0.01	2	3	0.01	2.5	3	0.01	3	3	0.01
$\alpha_1 = 1.0$ and $\alpha_2 = 3.0$														
β_1	β_2	$\hat{\rho}$	β_1	β_2	$\hat{\rho}$	β_1	β_2	$\hat{\rho}$	β_1	β_2	$\hat{\rho}$	β_1	β_2	$\hat{\rho}$
1	1	0.01	1.5	1	0.01	2	1	0.01	2.5	1	0.01	3	1	0.00
1	1.5	0.01	1.5	1.5	0.01	2	1.5	0.01	2.5	1.5	0.01	3	1.5	0.00
1	2.0	0.01	1.5	2	0.01	2	2	0.01	2.5	2	0.00	3	2	0.00
1	2.5	0.01	1.5	2.5	0.01	2	2.5	0.00	2.5	2.5	0.00	3	2.5	0.00
1	3	0.00	1.5	3	0.00	2	3	0.00	2.5	3	0.00	3	3	0.00

clinical setting, most of the biomarkers do not take any infinite values. The limited range of maximum and minimum values of biomarkers can be standardized into 0 to 1 scale range. In this scenario, Beta distribution is suitable enough to handle with biomarker than normal distribution as the parametric choice. It is also difficult to get a large number of the sample size of multiple biomarker measurements in any clinical setting due to cost and ethical issues. Hence, it is appropriate to assume Beta distribution than the normal distribution.

The characteristics of the proposed method have also been observed through changes of different correlation measurement and changes of parametric values. The MCMC iteration method is used to explore the approximate value of $\widehat{\text{AUC}}$. The proposed method is illustrated with a real-life example of prostate cancer data set.

The bivariate beta ROC model is applied to obtain the relation between FPR and TPR. The optimum value of bivariate Beta distribution on example data is obtained through estimates of AUC on combine accuracy of the two biomarkers together. The simulation studies have been conducted to obtain the estimates of different correlation coefficient starting from 0 to 0.9. It is assumed that individual is diseased when the values of Y_1 and Y_2 are large enough. The bivariate ROC curve for each biomarker plotted separately and it is concluded that the bivariate beta ROC model can be applied once the data fit the bivariate Beta distribution.

TABLE 7: Estimated parameter of δ_1C of bivariate beta for different shape and scale parameters.

$\alpha_1 = 1.0$ and $\alpha_2 = 1.0$														
β_1	β_2	$\hat{\delta}_1$	β_1	β_2	$\hat{\delta}_1$	β_1	β_2	$\hat{\delta}_1$	β_1	β_2	$\hat{\delta}_1$	β_1	β_2	$\hat{\delta}_1$
1	1	0.02	1.5	1	0.01	2	1	0.01	2.5	1	0.02	3	1	0.01
1	1.5	0.01	1.5	1.5	0.01	2	1.5	0.01	2.5	1.5	0.02	3	1.5	0.01
1	2.0	0.01	1.5	2	0.01	2	2	0.01	2.5	2	0.01	3	2	0.01
1	2.5	0.01	1.5	2.5	0.01	2	2.5	0.01	2.5	2.5	0.01	3	2.5	0.01
1	3	0.01	1.5	3	0.01	2	3	0.01	2.5	3	0.01	3	3	0.01
$\alpha_1 = 1.0$ and $\alpha_2 = 1.5$														
β_1	β_2	$\hat{\delta}_1$	β_1	β_2	$\hat{\delta}_1$	β_1	β_2	$\hat{\delta}_1$	β_1	β_2	$\hat{\delta}_1$	β_1	β_2	$\hat{\delta}_1$
1	1	0.01	1.5	1	0.01	2	1	0.01	2.5	1	0.01	3	1	0.01
1	1.5	0.01	1.5	1.5	0.01	2	1.5	0.01	2.5	1.5	0.01	3	1.5	0.01
1	2.0	0.01	1.5	2	0.01	2	2	0.01	2.5	2	0.01	3	2	0.01
1	2.5	0.01	1.5	2.5	0.01	2	2.5	0.01	2.5	2.5	0.01	3	2.5	0.01
1	3	0.01	1.5	3	0.01	2	3	0.01	2.5	3	0.01	3	3	0.01
$\alpha_1 = 1.0$ and $\alpha_2 = 2.0$														
β_1	β_2	$\hat{\delta}_1$	β_1	β_2	$\hat{\delta}_1$	β_1	β_2	$\hat{\delta}_1$	β_1	β_2	$\hat{\delta}_1$	β_1	β_2	$\hat{\delta}_1$
1	1	0.01	1.5	1	0.01	2	1	0.01	2.5	1	0.01	3	1	0.01
1	1.5	0.01	1.5	1.5	0.01	2	1.5	0.01	2.5	1.5	0.01	3	1.5	0.01
1	2.0	0.01	1.5	2	0.01	2	2	0.01	2.5	2	0.01	3	2	0.01
1	2.5	0.01	1.5	2.5	0.01	2	2.5	0.01	2.5	2.5	0.01	3	2.5	0.01
1	3	0.01	1.5	3	0.01	2	3	0.01	2.5	3	0.01	3	3	0.01
$\alpha_1 = 1.0$ and $\alpha_2 = 2.5$														
β_1	β_2	$\hat{\delta}_1$	β_1	β_2	$\hat{\delta}_1$	β_1	β_2	$\hat{\delta}_1$	β_1	β_2	$\hat{\delta}_1$	β_1	β_2	$\hat{\delta}_1$
1	1	0.01	1.5	1	0.01	2	1	0.01	2.5	1	0.01	3	1	0.01
1	1.5	0.01	1.5	1.5	0.01	2	1.5	0.01	2.5	1.5	0.01	3	1.5	0.01
1	2.0	0.01	1.5	2	0.01	2	2	0.01	2.5	2	0.01	3	2	0.01
1	2.5	0.01	1.5	2.5	0.01	2	2.5	0.01	2.5	2.5	0.01	3	2.5	0.01
1	3	0.01	1.5	3	0.01	2	3	0.01	2.5	3	0.01	3	3	0.01
$\alpha_1 = 1.0$ and $\alpha_2 = 3.0$														
β_1	β_2	$\hat{\delta}_1$	β_1	β_2	$\hat{\delta}_1$	β_1	β_2	$\hat{\delta}_1$	β_1	β_2	$\hat{\delta}_1$	β_1	β_2	$\hat{\delta}_1$
1	1	0.01	1.5	1	0.01	2	1	0.01	2.5	1	0.01	3	1	0.01
1	1.5	0.01	1.5	1.5	0.01	2	1.5	0.01	2.5	1.5	0.01	3	1.5	0.01
1	2.0	0.01	1.5	2	0.01	2	2	0.01	2.5	2	0.01	3	2	0.01
1	2.5	0.01	1.5	2.5	0.01	2	2.5	0.01	2.5	2.5	0.01	3	2.5	0.01
1	3	0.01	1.5	3	0.01	2	3	0.01	2.5	3	0.01	3	3	0.01

TABLE 8: Estimated parameter of δ_1C of bivariate beta for different shape and scale parameters.

$\alpha_1 = 1.0$ and $\alpha_2 = 1.0$														
β_1	β_2	$\hat{\delta}_2$	β_1	β_2	$\hat{\delta}_2$	β_1	β_2	$\hat{\delta}_2$	β_1	β_2	$\hat{\delta}_2$	β_1	β_2	$\hat{\delta}_2$
1	1	0.02	1.5	1	0.01	2	1	0.01	2.5	1	0.01	3	1	0.01
1	1.5	0.02	1.5	1.5	0.01	2	1.5	0.01	2.5	1.5	0.01	3	1.5	0.01
1	2.0	0.02	1.5	2	0.01	2	2	0.01	2.5	2	0.01	3	2	0.01
1	2.5	0.02	1.5	2.5	0.01	2	2.5	0.01	2.5	2.5	0.01	3	2.5	0.01
1	3	0.02	1.5	3	0.01	2	3	0.01	2.5	3	0.01	3	3	0.01
$\alpha_1 = 1.0$ and $\alpha_2 = 1.5$														
β_1	β_2	$\hat{\delta}_1$	β_1	β_2	$\hat{\delta}_1$	β_1	β_2	$\hat{\delta}_1$	β_1	β_2	$\hat{\delta}_1$	β_1	β_2	$\hat{\delta}_1$
1	1	0.02	1.5	1	0.01	2	1	0.01	2.5	1	0.01	3	1	0.01
1	1.5	0.02	1.5	1.5	0.01	2	1.5	0.01	2.5	1.5	0.01	3	1.5	0.01
1	2.0	0.02	1.5	2	0.01	2	2	0.01	2.5	2	0.01	3	2	0.01
1	2.5	0.02	1.5	2.5	0.01	2	2.5	0.01	2.5	2.5	0.01	3	2.5	0.01
1	3	0.02	1.5	3	0.01	2	3	0.01	2.5	3	0.01	3	3	0.01
$\alpha_1 = 1.0$ and $\alpha_2 = 2.0$														
β_1	β_2	$\hat{\delta}_1$	β_1	β_2	$\hat{\delta}_1$	β_1	β_2	$\hat{\delta}_1$	β_1	β_2	$\hat{\delta}_1$	β_1	β_2	$\hat{\delta}_1$
1	1	0.02	1.5	1	0.01	2	1	0.01	2.5	1	0.01	3	1	0.01
1	1.5	0.02	1.5	1.5	0.01	2	1.5	0.01	2.5	1.5	0.01	3	1.5	0.01
1	2.0	0.02	1.5	2	0.01	2	2	0.01	2.5	2	0.01	3	2	0.01
1	2.5	0.02	1.5	2.5	0.02	2	2.5	0.01	2.5	2.5	0.01	3	2.5	0.01
1	3	0.02	1.5	3	0.02	2	3	0.01	2.5	3	0.01	3	3	0.01
$\alpha_1 = 1.0$ and $\alpha_2 = 2.5$														
β_1	β_2	$\hat{\delta}_1$	β_1	β_2	$\hat{\delta}_1$	β_1	β_2	$\hat{\delta}_1$	β_1	β_2	$\hat{\delta}_1$	β_1	β_2	$\hat{\delta}_1$
1	1	0.02	1.5	1	0.01	2	1	0.01	2.5	1	0.01	3	1	0.01
1	1.5	0.02	1.5	1.5	0.01	2	1.5	0.01	2.5	1.5	0.01	3	1.5	0.01
1	2.0	0.02	1.5	2	0.02	2	2	0.01	2.5	2	0.01	3	2	0.01
1	2.5	0.02	1.5	2.5	0.02	2	2.5	0.01	2.5	2.5	0.01	3	2.5	0.01
1	3	0.02	1.5	3	0.02	2	3	0.01	2.5	3	0.01	3	3	0.01
$\alpha_1 = 1.0$ and $\alpha_2 = 3.0$														
β_1	β_2	$\hat{\delta}_1$	β_1	β_2	$\hat{\delta}_1$	β_1	β_2	$\hat{\delta}_1$	β_1	β_2	$\hat{\delta}_1$	β_1	β_2	$\hat{\delta}_1$
1	1	0.02	1.5	1	0.01	2	1	0.01	2.5	1	0.01	3	1	0.01
1	1.5	0.02	1.5	1.5	0.01	2	1.5	0.01	2.5	1.5	0.01	3	1.5	0.01
1	2.0	0.02	1.5	2	0.02	2	2	0.01	2.5	2	0.01	3	2	0.01
1	2.5	0.02	1.5	2.5	0.02	2	2.5	0.01	2.5	2.5	0.01	3	2.5	0.01
1	3	0.02	1.5	3	0.02	2	3	0.01	2.5	3	0.01	3	3	0.01

TABLE 9: Posterior estimated parameters, $\widehat{AUC}$,SD($\widehat{AUC}$) and 95% credible interval for AUC of bivariate beta ROC curve for different shape and scale parameters.

Initial Value	Parameter	Mean	SD	MC Error	2.5%	*Median*	97.5%
$\alpha_1 = 1, \alpha_2 = 1, \beta_1 = 1, \beta_2 = 2$	AUC	0.9882	0.0841	0.0027	0.9585	0.9721	0.9980
$\alpha_1 = 1, \alpha_2 = 2, \beta_1 = 1, \beta_2 = 2$	AUC	0.9779	0.0256	0.0080	0.9417	0.97277	0.9974
$\alpha_1 = 1, \alpha_2 = 2, \beta_1 = 2, \beta_2 = 1$	AUC	0.9645	0.0189	0.0043	0.9249	0.9563	0.9741
$\alpha_1 = 2, \alpha_2 = 1, \beta_1 = 1, \beta_2 = 2$	AUC	0.9616	0.0190	0.0056	0.9308	0.9532	0.9904
$\alpha_1 = 2, \alpha_2 = 1, \beta_1 = 2, \beta_2 = 1$	AUC	0.7681	0.0213	0.0065	0.7412	0.7342	0.8065
$\alpha_1 = 1, \alpha_2 = 1, \beta_1 = 2, \beta_2 = 2$	AUC	0.8356	0.0314	0.0086	0.8122	0.7855	0.8690

Conflict of Interest

Authors declared no conflict of interest or financial support.

Authorship Contributions

Idea/Concept: *Atanu Bhattacharjee,Gajendra Vishwakarma*
Design: *Atanu Bhattacharjee,Gajendra Vishwakarma*
Control/Supervision: *Atanu Bhattacharjee,Gajendra Vishwakarma*
Data Collection and/or Processing: *Atanu Bhattacharjee,Gajendra Vishwakarma*
Analysis and/or Interpretation: *Atanu Bhattacharjee,Gajendra Vishwakarma, Abin Thomas*

Literature Review: *Atanu Bhattacharjee*
Writing the Article: *Atanu Bhattacharjee,Gajendra Vishwakarma, Abin Thomas*
Critical Review: *Atanu Bhattacharjee,Gajendra Vishwakarma*
References and Fundings: *Atanu Bhattacharjee,Gajendra Vishwakarma*
Materials: *Atanu Bhattacharjee,Gajendra Vishwakarma, Abin Thomas*

REFERENCES

1. Cabral HJ. Multiple comparisons procedures. Circulation 2008;117(5):698-701.
2. Arnold BC, Castillo E, Sarabia JM. Conditionally specified distributions: an introduction (with comments and a rejoinder by the authors). Statist Sci 2001;16(3):249-74.
3. Arnold BC, Ng, Hon KT. Flexible bivariate beta distributions. J Multivar Anal 2011;102(8):1194-202.
4. Arnold BC, Castillo E, Sarabia JM. Families of multivariate distributions involving the Rosenblatt construction. J Am Stat Assoc 2006;101(476):1652-62.
5. Sarabia JM, Déniz EG. Construction of multivariate distributions: a review of some recent results. SORT 2008;32(1):3-36.
6. Gupta AK, Wong CF. On three and five parameter bivariate beta distributions. Metrika 1985;32(1):85-91.
7. Balakrishnan N, Lai CD. Continuous Bivariate Distributions. 2nd ed. New York: Springer Science & Business Media; 2009. p.1-17.
8. Jones MC. Multivariate t and beta distributions associated with the multivariate F distribution. Metrika 2002;54(3):215-31.
9. Nadarajah S, Kotz S. Some bivariate beta distributions. Statistics 2005;39(5):457-66.
10. Nadarajah S. A bivariate distribution with gamma and beta marginals with application to drought data. J Appl Stat 2009;36(3):277-301.
11. Wright S. The distribution of gene frequencies in populations. Proc Natl Acad Sci U S A 1937;23(6):307-20.
12. Gianola D, Manfredi E, Simianer H. On measures of association among genetic variables. Animal Genetics 2012;43(s1):19-35.
13. Nadarajah SA. A new bivariate beta distribution with application to drought data. Metron 2007;65(2):153-74.
14. Bibby BM, Væth M. The two-dimensional beta binomial distribution. Statistics & Probability Letters 2011;81(7): 884-91.
15. Wang M, Rennolls K. Bivariate distribution modeling with tree diameter and height data. Forest Science 2007;53(1):16-24.
16. Adell N, Puig P, Rojas-Olivares A, Caja G, Carnfie S, Salama Ahmed AK. A bivariate model for retinal image identification in lambs. Comput Electron Agric 2012;87:108-12.
17. Danaher PJ, Hardie BGS. Bacon with your eggs? Applications of a new bivariate beta-binomial distribution. Am Stat 2005;59(4):282-6.
18. Eugene N, Lee C, Famoye F. Beta-normal distribution and its applications. Commun Stat Theory Methods 2002;31(4):497-512.
19. Jones MC, Larsen PV. Multivariate distributions with support above the diagonal. Biometrika 2004;91(4):975-86.
20. Massman WJ. Foliage distribution in old-growth coniferous tree canopies. Can J For Res 1982;12(1):10-7.
21. Pitkänen S. The use of diversity indices to assess the diversity of vegetation in managed boreal forests. Forest Ecology & Management 1998;112(1):121-37.
22. Collett D. Statistical inference for binary data. Modelling Binary Data. 2nd ed. Boca Raton: CRC Press; 2002. p.19-42.
23. L'Ecuyer P. Non-uniform random variate generations. In: Lovric M, ed. International Encyclopedia of Statistical Science. Berlin, New York: Springer; 2011. p.991-5.
24. Johnson NL, Kotz S, Balakrishnan N. Discrete Multivariate Distributions. New York: Wiley; 1997. p.165.
25. Plackett RL. A class of bivariate distributions. J Am Stat Assoc 1965;60(312):516-22.
26. Lloyd CJ. Confidence intervals from the difference between two correlated proportions. J Am Stat Assoc 1990;85(412):1154-8.
27. Jones MC. Multivariate t and beta distributions associated with the multivariate F distribution. Metrika 2002;54(3):215-31.
28. Michael JR, Schucany WR. The mixture approach for simulating bivariate distributions with specific correlations. Am Stat 2002;56(1):48-54.
29. Olkin I, Liu R. A bivariate beta distribution. Statistics & Probability Letters 2003;62(4):407-12.
30. Libby DL, Novick MR. Multivariate generalized beta distributions with applications to utility assessment. J Educ Behav Stat 1982;7(4):271-94.
31. Casella G, Berger RL. In: Luc D, ed. Non-Uniform Random Variate Generation. New York: Springer; 1986. p.1-356.
32. El-Bassiouny AH, Jones MC. A bivariate F distribution with marginals on arbitrary numerator and denominator degrees of freedom, and related bivariate beta and t distributions. Stat Methods Appt 2009;18(4):465-81.
33. Park CG, Park T, Shin DW. A simple method for generating correlated binary variates. Am Stat 1996;50(4):306-10.
34. Johnson NL, Kotz S. Distributions in Statistics: Continuous Univariate Distributions. Vol. 1. New York: Houghton Miffin; 1970. p.300.
35. Steen M. An algorithm for generating positively correlated beta-distributed random variables with known marginal distributions and a specified correlations. Comput Stat Data An 2004;46:397-406.
36. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015;43(7):e47.
37. Wang MC, Li S. Bivariate makers measurements and ROC analysis. Biometrics 2012;66(4):1207-18.
38. Ma S, Huang J. Combining multiple markers for classification using ROC. Biometrics 2007;63(3):751-7.
39. Apostolakis G, Moieni P. The foundations of models of dependence in probabilistic safety assessment. Reliability Engineering 1987;18(3):177-95.
40. Xie M, Liu RY, Damaraju CV, Olson WH. Incorporating external information in analyses of clinical trials with binary outcomes. Ann Appl Stat 2013;7(1):342-68.
41. Oleson JJ. Bayesian credible intervals for binomial proportions in a single patient trial. Stat Methods Med Res 2010;19(6):559-74.