The Bi-Epsilon Skew Exponential Power (BIESEP) ROC Curve

Ahmad Flaih * Hassan Elsalloukh †

Abstract

A new model of the Receiver Operating Characteristic (ROC) curve, the Bi-Epsilon Skew Exponential Power (BIESEP) ROC curve, is proposed in this paper. This model is a generalization of the Epsilon Skew Binormal ROC curve. Elsalloukh (2004 and 2008) provided a flexible model, the Epsilon Skew Exponential Power (ESEP), which can be adopted to accommodate asymmetry and kurtosis (platykurtic or leptokurtic) tails. The ESEP model is an appropriate choice to increase the robustness of data analysis. We develop the binormal ROC curve with a diagnostic test outcome distributed according to the ESEP model. More specifically, we derive the BIESEP ROC accuracy function. Also, we consider the estimation of BIESEP ROC curve and accuracy of a diagnostic test.

KEY WORDS: Receiver Operating Characteristic, Bi-Epsilon Skew Exponential Power (BIESEP) ROC curve, Area Under the Curve.

1. Introduction

Medical diagnostic test is frequent in medicine practice since it plays an important role in discriminate between different health states,e.g. diseased and non-diseased. The Receiver Operating Characteristic (ROC) curve is appropriate and well accepted statistical tool for displaying the performance and accuracy of a medical diagnostic test in situations where there are two possible states, diseased / non-diseased , event/ non-event, or any binary outcome. Some standard methods to estimate the ROC curve and the related measures are parametric, non-parametric, and semiparametric methods. The parametric approach specifies a distribution for the diagnostic test outcomes, the non-parametric methods do not require any assumptions on either the density function of data or the function of the ROC curve. The semi-parametric methods assume the ROC curve as a smooth function, and come with fewer assumptions than the parametric methods.

In this research, we suppose that the outcome measurement Y of a medical diagnostic test results is continuous and distributed according to the Epsilon Skew Eponential Power (ESEP) distribution family proposed by Elsalloukh (2004 and 2005). We then develop an ROC curve when these outcomes are distributed as ESEP.

*UALR, Department of Applied Science, 2801 South University Avenue, Little Rock, AR 72204

†UALR, Department of Mathematics and Statistics, 2801 South University Avenue, Little Rock, AR 72204

2. The Receiver Operating Characteristic (ROC) Curve and the Area Under Curve (AUC)

In this section, we give an overview for the receiver operating characteristic ROC curve as a statistical decision tool. Faraggi and Reiser (2002) studied some nonparametric and parametric methods to estimate and compare the area under the ROC curve. Betinec (2008) developed the ROC curve based on the exponential distribution as a distribution for the diagnostic test measurments. Zou et al. (2003) discused two parametric models, bi-normal and bi-weibull models, and developed agoodness of fit test for the ROC curve. The medical diagnostic test with a continuous outcome Y distributed according to a normal distribution for detecting the disease assuming t as the threshold (cutoff) value of Y is called the Binormal ROC curve Dorfman and Alf (1968). The binormal ROC curve with a given threshold t is commonly assessed using the probabilities that correctly classify outcomes, which are called True Positive Rate (TPR) and False Positive Rate (FPR), defined as, respectively

$$TPR = P(Y_D > t/diseased) = 1 - \Phi(\frac{t - \theta_D}{\sigma_D}) \tag{1}$$

$$FPR = P(Y_{\ddot{D}} > t/non - diseased) = 1 - \Phi(\frac{t - \theta_{\ddot{D}}}{\sigma_{\ddot{D}}}), \tag{2}$$

where $Y_D \sim N(\theta_D, \sigma_D^2)$ and $Y_{\ddot{D}} \sim N(\theta_{\ddot{D}}, \sigma_{\ddot{D}}^2)$ are diagnostic test outcomes for detecting whether a subject is diseased (D) and /or nondiseased ($\ddot{D}$), and $\Phi(\cdot)$ is the standard normal cdf. Suppose $\theta_D > \theta_{\ddot{D}}$, then we can define the ROC curve as

$$ROC(t) = (FPR(t), TPR(t)); t \in \Re. \tag{3}$$

Hence, the Binormal ROC curve equation is defined by

$$h(x, \theta) = 1 - \Phi[-a + b\Phi^{-1}(1 - x)]; 0 \leq x \leq 1,$$

where

$$a = \frac{\theta_D - \theta_{\ddot{D}}}{\sigma_D}, b = \frac{\sigma_{\ddot{D}}}{\sigma_D}.$$

The most common used summary index of the performance of a diagnostic test based on an ROC curve, Krzanowski and Hand (2009), is the Area Under the ROC Curve, denoted by (AUC), that is

$$AUC = P(Y_D \geq Y_{\ddot{D}}) = \Phi(\frac{\theta_D - \theta_{\ddot{D}}}{\sqrt{\sigma_{\ddot{D}}^2 + \sigma_D^2}}).$$

Moreover, AUC index is a popular summary measure of diagnostic test accuracy based on an ROC curve, Bamber (1975). Mashtare Jr. and Huston (2009) considered the epsilon skew binormal ROC curve, derived TPR and FPR equations, estimated the area under the ROC curve, and discussed its application in biomedical.

3. The Epsilon Skew Exponential Power (ESEP) distribution

Elsalloukh (2004 and 2005) introduced the Epsilon-Skew Exponential Power (ESEP) distribution that can accommodate heavy-tailed (Leptokurtic) and skewed data. The ESEP distribution is attractive and flexible because it allows continuous variation from normality to non-normality and nested with many models especially with the normal distribution, that is, the ESEP includes the normal distribution as a special case and hence is a "robust model". The ESEP density is denoted by $ESEP(\theta,\sigma,\alpha,\epsilon)$ and defined by

$$f(y)=\frac{\alpha}{2\sigma\sqrt{2}\Gamma(\frac{1}{\alpha})}\begin{cases}exp[-(\frac{y-\theta}{\sqrt{2}\sigma(1-\epsilon)})^{\alpha}]; & y\geq\theta\\ exp[-(\frac{\theta-y}{\sqrt{2}\sigma(1+\epsilon)})^{\alpha}]; & y<\theta,\end{cases}\tag{4}$$

where $-1<\epsilon<1$ is the skewness parameter, $\theta\in\Re$ is the location parameter, $\sigma>0$ is the scale parameter, and $\alpha\in\Re$ is the shape parameter. Moreover, the density function (4) is known as the Epsilon Skew Exponential Power of order α.

The cumulative distribution function of the $ESEP(\theta,\sigma,\alpha,\epsilon)$

$$F(y)=\begin{cases}1-\frac{(1-\epsilon)}{2\Gamma(\frac{1}{\alpha})}\Gamma(\frac{1}{\alpha},g(y)); & y\geq\theta\\ \frac{(1+\epsilon)}{2\Gamma(\frac{1}{\alpha})}\Gamma(\frac{1}{\alpha},h(y)); & y<\theta,\end{cases}\tag{5}$$

where, $\Gamma(\cdot,\cdot)$ is the incomplete gamma function, $g(y)=[\frac{y-\theta}{2^{\frac{1}{2}}(1-\epsilon)\sigma}]^{\alpha}$ and $h(y)=[\frac{\theta-y}{2^{\frac{1}{2}}(1+\epsilon)\sigma}]^{\alpha}$. The quantile function of Y is

$$F^{-1}_{ESEP}(v/\epsilon,\alpha)=\begin{cases}\theta-2^{\frac{1}{2}}\sigma(1+\epsilon)\\ [G^{-1}(\frac{v}{1+\epsilon}2\Gamma(\frac{1}{\alpha});\frac{1}{\alpha})]^{\frac{1}{\alpha}}; & 0<v<\frac{1+\epsilon}{2}\\ \theta+2^{\frac{1}{2}}\sigma(1-\epsilon)\\ [G^{-1}(\frac{1-v}{1-\epsilon}2\Gamma(\frac{1}{\alpha});\frac{1}{\alpha})]^{\frac{1}{\alpha}}; & \frac{1+\epsilon}{2}\leq v<1,\end{cases}\tag{6}$$

where $G^{-1}(\cdot)$ is the inverse function of the gamma cdf $G(\cdot)$, and

$$G(y,\gamma)=(\Gamma(\gamma))^{-1}\int_0^y z^{\gamma-1}exp(-z)dz,$$

4. The Bi-Epsilon Skew Exponential Power (BIESEP) ROC Curve

Consider the receiver operating characteristic ROC curve as defined in (3), the TPR and FPR based on the ESEP family are

$$TPR=1-P(Y_D\leq t)=1-F(\frac{t-\theta_D}{\sigma_D}),\tag{7}$$

$$FPR=1-P(Y_{\ddot{D}}\leq t)=1-F(\frac{t-\theta_{\ddot{D}}}{\sigma_{\ddot{D}}}).\tag{8}$$

Let $Y_{\ddot{D}}\sim ESEP(\theta_{\ddot{D}},\sigma_{\ddot{D}},\alpha_{\ddot{D}},\epsilon_{\ddot{D}})$, and $Y_D\sim ESEP(\theta_D,\sigma_D,\alpha_D,\epsilon_D)$ be the diagnostic test outcome for detecting whether a subject is diseases (D) or non-diseased $(\ddot{D})$. Let $F_{\ddot{D}}(\cdot)$ and $F_D(\cdot)$ be the cdfs for the standard ESEP, and suppose $\theta_D>\theta_{\ddot{D}}$.

proposition 1. *If $Y_{\ddot{D}}\sim ESEP(\theta_{\ddot{D}},\sigma_{\ddot{D}},\alpha_{\ddot{D}},\epsilon_{\ddot{D}})$, and $Y_D\sim ESEP(\theta_D,\sigma_D,\alpha_D,\epsilon_D)$ denote non-diseased and diseased test results, assuming $\theta_D>\theta_{\ddot{D}}$, then*

$$ROC(s,\boldsymbol{\lambda})=1-F_D[-\alpha+\beta F^{-1}_{\ddot{D}}(1-s)];s\in(0,1),\tag{9}$$

where

$$\alpha=\frac{\theta_D-\theta_{\ddot{D}}}{\sigma_D},\beta=\frac{\sigma_{\ddot{D}}}{\sigma_D},\boldsymbol{\lambda}=(\theta_D,\sigma_D,\alpha_D,\epsilon_D,\theta_{\ddot{D}},\sigma_{\ddot{D}},\alpha_{\ddot{D}},\epsilon_{\ddot{D}})$$

and $F^{-1}(\cdot)$ and $F(\cdot)$ the quantile function and the distribution function of the ESEP, respectively.

Proof. It is clear from (8), the cutoff can be expressed as

$$t = \theta_{\ddot{D}} + \sigma_{\ddot{D}} F_{\ddot{D}}^{-1}(1 - FPR). \tag{10}$$

By substituting (10) in (7), we have

$$\begin{aligned} TPR &= 1 - F(\frac{\theta_{\ddot{D}} + \sigma_{\ddot{D}} F_{\ddot{D}}^{-1}(1 - FPR) - \theta_D}{\sigma_D}) \\ &= 1 - F[-\frac{(\theta_D - \theta_{\ddot{D}})}{\sigma_D} + \frac{\sigma_{\ddot{D}}}{\sigma_D} F_{\ddot{D}}^{-1}(1 - FPR)]. \end{aligned} \tag{11}$$

The more convenient expression for equation (11) with the parameters α and β is

$$ROC(s) = 1 - F_D[-\alpha + \beta F_{\ddot{D}}^{-1}(1 - s)]; s \in (0, 1),$$

where $\alpha = \frac{\theta_D - \theta_{\ddot{D}}}{\sigma_D}, \beta = \frac{\sigma_{\ddot{D}}}{\sigma_D}$ are the parameters of the BIESEP ROC curve. □

proposition 2. *The MLE $\hat{\boldsymbol{\lambda}}$ of $\boldsymbol{\lambda}$ is asymptotically normal, i.e.,*

$$\sqrt{n}(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}) \to^d N(0, \Sigma), \tag{12}$$

where Σ is the variance-covariance matrix.

Proof. Straightforward from Theorem 4.1. in Elsalloukh (2004), and as $n \to \infty$, with variance-covarince matrix of the estimated parameters of two i.i.d. ESESP random variables,

$$\Sigma = \left[\begin{array}{c|c} I_{\ddot{D}}^{-1} & 0 \\ \hline 0 & I_D^{-1} \end{array}\right] \tag{13}$$

where I is the Fisher information matrix as defined in Theorem 4.1 in section 4.3 Elsalloukh (2004). Hence, we have (12). □

proposition 3. *Let Y_D and $Y_{\ddot{D}}$ denote any binary continuous random variables, then*

$$\begin{aligned} \widehat{AUC(s, \boldsymbol{\lambda})} &\simeq \int_0^1 ROC(s, \hat{\boldsymbol{\lambda}}) ds \\ &\simeq \frac{1}{m}[\frac{1}{2} g(0; \hat{\boldsymbol{\lambda}}) + \frac{1}{2} g(1; \hat{\boldsymbol{\lambda}}) + \sum_{i=2}^{m} g(\frac{i-1}{m}; \hat{\boldsymbol{\lambda}})] \\ &\simeq \frac{1}{m}[\sum_{i=0}^{m} g_i(\hat{\boldsymbol{\lambda}})], \end{aligned} \tag{14}$$

where m is the number of intervals, each of size $\frac{1}{m}$ and $ROC(s, \hat{\boldsymbol{\lambda}}) = g(\cdot)$.

Proof. Using the fact that the trapezoidal rule is to break up the range from $a = 0$ to $b = 1$ into m smaller intervals, each of size $w = \frac{b-a}{m}$, and the areas of all the trapezoids:

$$\int_a^b f(y) dy \simeq \frac{w}{2}[f(a) + f(b) + 2\sum_{i=2}^{m} f(a + (i-1)w)], \tag{15}$$

using $a = 0$,$b = 1$, and $w = \frac{1}{m}$, the approximate area of the BIESEP ROC curve equation (9) becomes (14). □

proposition 4. *Let $\widehat{AUC(s,\boldsymbol{\lambda}})$ as defined in (14) with $\hat{\boldsymbol{\lambda}} \sim AN(\boldsymbol{\lambda}, \Sigma_m)$, then for large samples $\widehat{AUC(s,\boldsymbol{\lambda}})$ is $AN(AUC(s,\boldsymbol{\lambda}), Var(\widehat{AUC(s,\boldsymbol{\lambda}})))$, where*

$$Var(\widehat{AUC(s,\boldsymbol{\lambda}})) = \frac{1}{m^2}[\sum_{i=0}^{m} g_i(\hat{\boldsymbol{\lambda}}) + 2\sum_{i=0}^{m}\sum_{j>i}^{m} \sigma_{ij}(g_i(\hat{\boldsymbol{\lambda}}), g_j(\hat{\boldsymbol{\lambda}}))]. \tag{16}$$

Proof. Using the fact, Corollary of Theorem A in Serfling (1980),

$$g(X_n) \sim AN(g(\mu), \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{k} \sigma_{ij}\frac{\partial^2 g}{\partial x_i x_j}|_{x=\mu}),$$

where

$$X_n \sim AN(\mu, n^{-1}\Sigma),$$

and Σ is a covariance matrix. □

5. Area Under the BIESEP ROC Curve Estimation

The area under the BIESEP ROC curve has been defined in section 4. as

$$\widehat{AUC(s,\boldsymbol{\lambda}}) \simeq \int_0^1 ROC(s,\hat{\boldsymbol{\lambda}})ds.$$

To estimate this quantity from sample data, we can estimate the ROC curve by fitting a smooth curve assuming each of the two populations (event / non-event) have been distributed by ESEP, and then obtain $\widehat{AUC(s,\boldsymbol{\lambda}})$ numerically by using Trapezoidal rule. In terms of the population parameters, $\boldsymbol{\lambda} = (\theta_D, \sigma_D, \alpha_D, \epsilon_D, \theta_{\ddot{D}}, \sigma_{\ddot{D}}, \alpha_{\ddot{D}}, \epsilon_{\ddot{D}})$, the $\widehat{AUC(s,\boldsymbol{\lambda}})$ can be approximated by

$$\widehat{AUC(s,\boldsymbol{\lambda}}) = \frac{1}{m}[\frac{1}{2} + \sum_{i=2}^{m-1} g(\frac{i-1}{m};\hat{\boldsymbol{\lambda}})], \tag{17}$$

where m is the number of intervals in the domain of the integration [a, b].

Lemma 5.1. *Let $g(s,\hat{\boldsymbol{\lambda}})$ be $(m-1)\times 1$ vector of the $g(\frac{i-1}{m},\hat{\boldsymbol{\lambda}})$,$i = 1, ..., m-1$, then as m approaches ∞,*

$$g(s,\hat{\boldsymbol{\lambda}}) \sim AN(g(s,\boldsymbol{\lambda}), \frac{1}{m}A\Sigma_m A'), \tag{18}$$

where A is the 8×1 vector of partial derivatives of the $g_i(\hat{\boldsymbol{\lambda}})$, and Σ_m is as defined in (13).

Proof. By proposition (3), and directly from theorem A, Serfling (1980),

$$g(X_n) \sim AN(g(\mu), b_n^2 D\Sigma\, D')$$

where

$$X_n \sim AN(\mu, n^{-1}\Sigma)$$

Σ is a covariance matrix, and

$$D = [\frac{\partial g_i}{\partial x_j}|_{x=\mu}]_{m\times k}$$

then, we have (18). □

proposition 5. *As m approaches infinity,*

$$\widehat{AUC(s,}\boldsymbol{\lambda}) \sim AN(AUC, \frac{1}{m^2}\mathbf{1}'_{m-1}D\mathbf{1}_{m-1}).where$$

$\mathbf{1}_{m-1}$ *is a* $(m-1)\times 1$ *vector of* $\mathbf{1}$*'s, and* $D = \frac{1}{m}A\Sigma_m A'$ *is as defined in* lemma *5.1.*

Proof. Straightforward derivation from Theorem 5.1.2. in Mashtare Jr. and Huston (2009), Theorem A in Serfling (1980), and using the fact that $\widehat{AUC(s,}\boldsymbol{\lambda})$ ia a real function of $g(\cdot,\hat{\boldsymbol{\lambda}})$ with

$$\frac{\partial \widehat{AUC(s,}\boldsymbol{\lambda})}{\partial g(\frac{i-1}{m},\hat{\boldsymbol{\lambda}})} = \frac{1}{m}$$

□

REFERENCES

Bamber, D. (1975)," The area above the ordinal dominance graph and the area below the receiver operating graph," Journal of Mathematical Psychology, 12, 387-415.

Betinec, M. (2008), "Testing the difference of the ROC curves in biexponential model," Tatra Mt. Math, Publ.39,215223.

Dorfman, D.D. and Alf, E. (1968), " Maximum likelihood estimation of parameters of signal detection theory and determination of confidence intervalsrating method data," Journal of Mathematical Psychology,6, 487-496.

Elsalloukh, H. (2004), " The epsilon skew exponential power distribution."PhD. thesis, Baylor university, USA.

Elsalloukh, H., Guardiola, J. and Young, M.D. (2005), "The Epsilon-Skew Exponential Power Distribution Family," Far East Journal of Theoretical Statistics, 16, No. 1, 97-112.

Elsalloukh, H. (2008), "The Epsilon-Skew Laplace Distribution," The 2008 Proceedings of the American Statistical Association, Biometrics Section.

Faraggi, D. and Reiser, B. (2002), "Estimation of the area under the ROC curve," Statistics in Medicine ,21,3093-3106.

Krzanowski, W.J. and Hand, D.J. (2009), "ROC Curves for Continuous Data," Taylor and Francis Group.

Mashtare Jr.,T. and Huston, A. (2009), "Epsilon-skew-binormal receiver operating characteristic (ROC) curves,"(in press).

Serfling, R. (1980), "Approximation theorems of mathematical statistics," Wiley and Sons: New York, NY.