

ROC Curve Estimation Based On Local Smoothing

Peihua Qiu¹ and Chap Le²

School of Statistics¹, University of Minnesota, 313 Ford Hall, Minneapolis, MN 55455

School of Public Health², University of Minnesota, 303 Mayo, Minneapolis, MN 55455

Abstract

ROC curve is a graphical representation of the relationship between sensitivity and specificity of a diagnostic test. It is a popular tool for evaluating and comparing different diagnostic tests in medical sciences. In the literature, the ROC curve is often estimated empirically based on an empirical distribution function estimator and an empirical quantile function estimator. In this paper an alternative nonparametric procedure to estimate the ROC curve is suggested which is based on local smoothing techniques. Several numerical examples are presented to evaluate the performance of this procedure.

Key Words: Diagnostic tests; Empirical estimators; Incomplete Beta function; Kernel Distribution function estimator; Quantile estimators; ROC curves;

1 Introduction

Suppose that a diagnostic test is based on a continuous measurement T . A person is classified as positive or diseased if $T \leq t$ and as negative or non-diseased otherwise, where t is a cutpoint. Sensitivity of the diagnostic test is defined by the probability that a diseased person is correctly classified as diseased. That is, $sensitivity(t) = P(T \leq t | \text{a diseased person})$. Specificity of the test is defined by the probability that a non-diseased person is classified as non-diseased. That is, $specificity(t) = P(T > t | \text{a normal person})$. The curve of the pair $(sensitivity(t), 1 - specificity(t))$ when t ranges over all possible values is called the receiver operating characteristic (ROC) curve of the diagnostic test, and has become a more and more popular tool in medical sciences to evaluate and compare different diagnostic tests (see e.g., Campbell 1994; Le 1997). It is apparent that a ROC curve is a non-decreasing curve that joins points $(0, 0)$ and $(1, 1)$.

¹Author for correspondence. E-mail: qiu@stat.umn.edu

Let X denote measurement T for non-diseased population with distribution function $F(x)$, and Y for diseased population with distribution function $G(y)$. Then $sensitivity(t) = G(t)$, $1 - specificity(t) = F(t)$, and the ROC curve is a mapping $R(\cdot)$ from $F(t)$ to $G(t)$:

$$G(t) = R(F(t)), \text{ for } t \in (-\infty, \infty). \quad (1.1)$$

Suppose that $F^{-1}(\cdot)$ is the corresponding quantile function of the distribution function $F(x)$ defined in the conventional way. Then (1.1) is equivalent to:

$$R(s) = G(F^{-1}(s)), \text{ for } s \in [0, 1]. \quad (1.2)$$

In other words, the ROC function $R(\cdot)$ (which defines the ROC curve) is determined by the distribution function $G(\cdot)$ and the quantile function $F^{-1}(\cdot)$.

In the literature, many authors use the empirical procedure to fit ROC curves (e.g., DeLong, DeLong and Clarke-Pearson 1988; Hanley and McNeil 1982; Hin *et al.* 1997). Given two independent samples $\{X_i\}_{i=1}^n$ and $\{Y_j\}_{j=1}^m$ from the non-diseased and diseased populations, respectively, the empirical estimators of $G(\cdot)$ and $F^{-1}(\cdot)$ are defined by

$$\begin{aligned} \hat{G}_m(y) &:= \{\text{number of } Y_j\text{'s } \leq y\}/m, \\ \hat{F}_n^{-1}(s) &:= X_{([ns]+1)}, \end{aligned} \quad (1.3)$$

where $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ are the order statistics of $\{X_i\}_{i=1}^n$ and $[x]$ represents the integer part of x . (Please notice that different authors may use different versions of the empirical estimator of $F^{-1}(\cdot)$.) Then the empirical estimator of the ROC function is defined by

$$\hat{R}_E(s) := \hat{G}_m(\hat{F}_n^{-1}(s)), \text{ for } s \in [0, 1]. \quad (1.4)$$

Another popular ROC curve fitting procedure is based on the conventional binormal model which assumes that both the diseased and non-diseased populations are Gaussian or could be transformed to Gaussian by a single monotonic transformation (Dorfman and Alf 1969; Swets 1979). The related parameters in such model are often estimated by maximum likelihood estimation. To overcome the problems of “hooks” and “degeneracy” in ROC curves that are fitted by the binormal model, Metz and Pan (1999), Pan and Metz (1997) and several others suggested using a “proper” binormal model which used likelihood ratio of the two population distributions as a decision variable.

Recently Lloyd (1998) suggested a local smoothing estimator of the ROC curve. More specifically his ROC curve estimator can be obtained by plotting a kernel estimator of $G(\cdot)$ against a kernel estimator of $F(\cdot)$. He also suggested a transform of the ROC curve based on the idea of “local population separation” for displaying the differences between the diseased and non-diseased populations.

As mentioned above, the ROC curve is a combination of a distribution function and a quantile function. In the literature, nonparametric estimation of a distribution function and a quantile function has been well studied during the past thirty years. Some methods there could be applied to the ROC curve fitting, which is a main motivation of this research.

Our focus will be on estimating the entire ROC curve from noisy data. Various summary statistics such as the area under the curve (AUC) can be computed from the fitted ROC curves. Although the AUC statistic is the most commonly used summary statistic in applications, other summary statistics are also possible and it seems to us that there is no consensus on which statistic should be used for a specific application (see e.g., Zweig and Campbell 1993 for related discussions).

In the empirical ROC curve fitting procedure, both $\hat{G}_m(\cdot)$ and $\hat{F}_n^{-1}(\cdot)$ are simple to compute. But they may not be the best nonparametric estimators of $G(\cdot)$ and $F^{-1}(\cdot)$, respectively (see e.g., Cai and Roussas 1998; Dielman *et al.* 1994). Consequently, it can be expected that $\hat{R}_E(s)$ would not be the best estimator of $R(s)$ in general. In this paper, we suggest a nonparametric estimator of $R(s)$ which is based on a kernel distribution function estimator of $G(\cdot)$ and a local smoothing quantile function estimator of $F^{-1}(\cdot)$. This estimator is described in detail in Section 2. Numerical examples in Section 3 show that it works better than the empirical procedure in many cases.

2 An Alternative Nonparametric Procedure

As mentioned in Section 1, the empirical ROC curve estimator is based on an empirical estimator of a quantile function and an empirical estimator of a distribution function. In the literature, there is extensive discussion about quantile function estimation (e.g., Eubank 1985; Parrish 1990). The empirical quantile estimator $X_{([ns]+1)}$ and its variants use sample quantiles $X_{([ns])}$, $X_{([ns]+1)}$ or their weighted averages to estimate population quantile $F^{-1}(s)$. Several authors pointed out that these empirical estimators were not efficient in some cases because they ignored all the helpful

information in all other order statistics (other than $X_{([ns]}$ and $X_{([ns]+1)}$). After performing several large simulation studies, Dielman *et al.* (1994), Harrell and Davis (1982) and Parrish (1990), among several others, found that the following Harrell-Davis quantile estimator performed better than the empirical estimators in many situations:

$$\widehat{F}_{H-D}^{-1}(s) := \sum_{j=1}^n W_{n,j} X_{(j)}, \quad (2.1)$$

where

$$W_{n,j} := I_{j/n}\{s(n+1), (1-s)(n+1)\} - I_{(j-1)/n}\{s(n+1), (1-s)(n+1)\},$$

and $I_x\{a, b\}$ denotes the incomplete beta function defined by:

$$I_x\{a, b\} := \frac{1}{B(a, b)} \int_0^x t^{a-1} (1-t)^{b-1} dt, \quad (2.2)$$

in which $B(a, b)$ is the complete beta function. More specifically they showed that $\widehat{F}_{H-D}^{-1}(s)$ outperformed the other quantile estimators when $F(\cdot)$ was a symmetric short-tailed distribution function such as the normal distribution function. When $F(\cdot)$ was long-tailed or skewed, its performance was good when $s \in [.1, .9]$ but not ideal when s was close to 0 or 1.

Besides the empirical estimator defined in (1.3), another popular distribution function estimator is the kernel-type estimator defined by:

$$\widehat{G}_{ker}(y) := \frac{1}{m} \sum_{j=1}^m K\left(\frac{y - Y_j}{h}\right), \quad (2.3)$$

where $K(\cdot)$ is the distribution function of a positive kernel function $k(x)$ (i.e., $K(x) = \int_{-\infty}^x k(t) dt$) and h is a window width. Several authors (e.g., Falk 1983; Reiss 1981) proved that the empirical estimator $\widehat{G}_m(y)$ was asymptotically deficient comparing with $\widehat{G}_{ker}(y)$ in the following sense. Let $i(m)$ be the first $k = 1, 2, \dots$ for which the MSE of $\widehat{G}_k(y)$ does not exceed $MSE(\widehat{G}_{ker}(y))$. Then for a suitable choice of $K(\cdot)$ and h , $i(m) - m$ tends to infinity when $m \rightarrow \infty$. In other words, the empirical estimator requires substantially larger sample size to achieve the same efficiency of the kernel-type estimator.

Therefore a natural estimator of the ROC function is defined by:

$$\widehat{R}_N(s) := \widehat{G}_{ker}(\widehat{F}_{H-D}^{-1}(s)). \quad (2.4)$$

The incomplete beta function in (2.2) can be computed by a Fortran code provided by Majumder

and Bhattacharjee (1973). Altman and Leger (1995) suggested a formula for computing the asymptotically optimal window width:

$$\hat{h}_{opt} = (.25\hat{V}/\hat{B})^{1/3}m^{-1/3},$$

where

$$\hat{V} = 2 \int_{-\infty}^{\infty} xk(x)K(x) dx \cdot \frac{1}{m(m-1)} \sum_{j_1 \neq j_2} \alpha_\nu^{-1} k_\nu\left(\frac{Y_{j_1} - Y_{j_2}}{\alpha_\nu}\right),$$

$$\hat{B} = .25 \left(\int_{-\infty}^{\infty} x^2 k(x) dx \right)^2 \cdot \frac{1}{m^3 \alpha_b^4} \sum_{j_1=1}^m \sum_{j_2=1}^m \sum_{j_3=1}^m k'_b\left(\frac{Y_{j_1} - Y_{j_2}}{\alpha_b}\right) k'_b\left(\frac{Y_{j_1} - Y_{j_3}}{\alpha_b}\right),$$

k_ν is a kernel function, α_ν is its window width, k'_b is the derivative of a kernel function k_b , and α_b is the associated window width. Altman and Leger (1995) also provided some guidelines for choosing k_ν , k_b , α_ν and α_b .

As noticed by Zweig and Campbell (1993), one disadvantage of the empirical estimator (1.4) of the ROC function is its staircase appearance. If a specific value of (1-specificity) happens to be a jump point of the empirical ROC curve estimator, then the estimator of the corresponding specificity value is not well defined by this method. As an example, suppose that we have the following ordered sample from the non-diseased population: 1.0,1.1,1.2,1.3,1.4,2.1,2.2,2.3,2.4,2.5; and the following ordered sample from the diseased population: 1.1,1.2,1.3,1.4,1.5,1.6,1.7,1.8,1.9,2.0. Then the estimated ROC curves by our procedure (2.4) and the empirical procedure (1.4) are presented in Figure 2.1 by the solid curve and the dotted curve, respectively. It can be seen from the plot that the empirical ROC curve estimator has a staircase appearance and our procedure smoothes out all of the staircases.

The ROC curve estimator (2.4) is explicitly defined by a composite of a distribution function estimator and a quantile function estimator. For a given value of 1-specificity, the corresponding estimator of the sensitivity value can be easily computed. To estimate the sensitivity value at a given 1-specificity level by Lloyd's (1998) procedure, we first need to figure out the corresponding value of the diagnostic measurement T and then the sensitivity value can be obtained by using the mapping between $F(\cdot)$ and $G(\cdot)$ (cf. (1.1)), making its application a little complicated. We also notice that both the empirical procedure and Lloyd's (1998) procedure treat the non-diseased and diseased populations symmetrically while the procedure (2.4) does not. In other words, if the status of the two populations were interchanged, then the empirical procedure and Lloyd's (1998) procedure could provide a same answer while the current procedure would most probably lead to

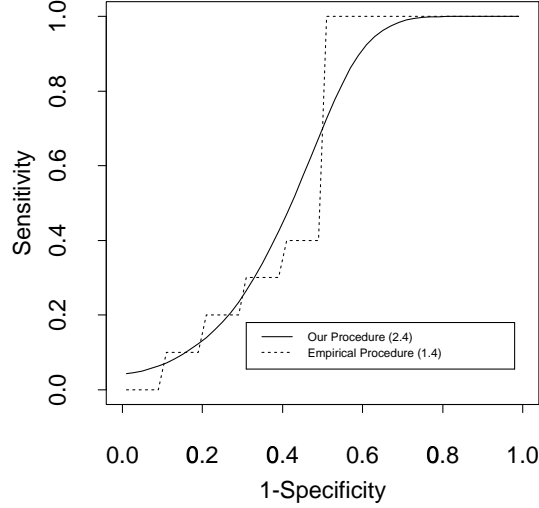


Figure 2.1: The solid and dotted curves represent the ROC curve estimators by our procedure (2.4) and the empirical procedure (1.4), respectively.

a different answer. It is also noteworthy that the ROC curve estimators by the current procedure and Lloyd's procedure are affected by strictly monotone transformations on the measurement T although such transformations would not affect the true ROC curve. The ROC curve estimator by the empirical procedure, however, is not affected by those transformations.

3 A Numerical Study

In the numerical examples presented in this section, we choose $k(x) \equiv k_\nu(x) \equiv k_b(x) = \frac{3}{4}(1 - x^2)I_{\{-1 \leq x \leq 1\}}$ and $\alpha_\nu = \alpha_b = m^{-0.3}$ which were suggested by Altman and Leger (1995). First, we assume that $X \sim N(2, 1), Y \sim N(0, 1)$ and $n = m = 20$. After generating two samples from the diseased and non-diseased populations, the empirical ROC curve estimator (1.4) and the nonparametric smoothing ROC curve estimator (2.4) are computed, respectively. They are plotted in Figure 3.1 along with the true ROC curve.

The two ROC curve fitting procedures are then compared by the following average squared error (ASE) criterion:

$$ASE(\hat{R}, R) := \frac{1}{M} \sum_{i=1}^M \left[\hat{R}\left(\frac{i-0.5}{M}\right) - R\left(\frac{i-0.5}{M}\right) \right]^2,$$

where \hat{R} denotes an estimator of R and M is a positive integer. The criterion $ASE(\hat{R}, R)$ averages M squared errors at M equally spaced (1-specificity) levels. In this section, M is fixed at 200.

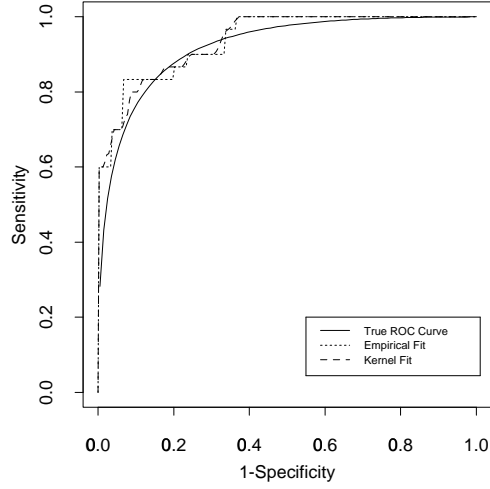


Figure 3.1: The fitted empirical ROC curve (dotted curve), the fitted nonparametric smoothing ROC curve (dashed curve) and the true ROC curve (solid curve).

Similarly, we define

$$ASE(\hat{F}^{-1}, F^{-1}) := \frac{1}{M} \sum_{i=1}^M \left[\hat{F}^{-1}\left(\frac{i-0.5}{M}\right) - F^{-1}\left(\frac{i-0.5}{M}\right) \right]^2$$

and

$$ASE(\hat{G}, G) := \frac{1}{M} \sum_{i=1}^M \left[\hat{G}\left(F^{-1}\left(\frac{i-0.5}{M}\right)\right) - G\left(F^{-1}\left(\frac{i-0.5}{M}\right)\right) \right]^2$$

to measure the accuracy of the quantile estimator \hat{F}^{-1} and the distribution function estimator \hat{G} .

We then let both n and m change their values among 5, 10, 20, 30, 40 and 50. For each combination of n and m , the simulation is repeated 1000 times. The averaged ASE values for both $ASE(\hat{F}_n^{-1}, F^{-1})$ and $ASE(\hat{F}_{H-D}^{-1}, F^{-1})$ are recorded and plotted in Figures 3.2(a) and 3.2(b). In Figure 3.2(c), $RASE := ASE(\hat{F}_n^{-1}, F^{-1})/ASE(\hat{F}_{H-D}^{-1}, F^{-1})$ is plotted. We can have two conclusions from these plots. (1) When the sample size n of the sample from the non-diseased population increases, the precision of both \hat{F}_n^{-1} and \hat{F}_{H-D}^{-1} increases. (2) \hat{F}_{H-D}^{-1} provides a more accurate estimator of F^{-1} and the benefit to use it is more obvious when n increases.

Similarly, $ASE(\hat{G}_m, G)$, $ASE(\hat{G}_{ker}, G)$ and $RASE := ASE(\hat{G}_m, G)/ASE(\hat{G}_{ker}, G)$ are plotted in Figures 3.2(d), 3.2(e) and 3.2(f), respectively. We can see that \hat{G}_{ker} performs better than \hat{G}_m but their relative difference is smaller when m gets larger.

In Figures 3.2(g), 3.2(h) and 3.2(i), $ASE(\hat{R}_E, R)$, $ASE(\hat{R}_N, R)$ and $RASE := ASE(\hat{R}_E, R)/ASE(\hat{R}_N, R)$ are presented, respectively. It can be seen that: (1) the two procedures both perform

better when m or n increases; (2) \widehat{R}_N provides a more accurate estimator of the ROC function than \widehat{R}_E ; and (3) the relative difference between \widehat{R}_N and \widehat{R}_E is quite large when either m or n is small and it tends to be stabilized when both m and n are large.

Intuitively, the degree of overlap between the two population distributions should be an important factor affecting the accuracy of the estimated ROC curve. In the above example, let us fix both n and m at 10. The mean of the non-diseased population distribution (denoted as a) is allowed to change its value among 0.1, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5 and 4.0. Clearly there is less overlap between the two population distributions when a is larger. The corresponding $ASE(\widehat{R}_E, R)$, $ASE(\widehat{R}_N, R)$ and $RASE := ASE(\widehat{R}_E, R)/ASE(\widehat{R}_N, R)$ values are plotted in Figures 3.3(a), 3.3(b) and 3.3(c), respectively. It can be seen from the plots that (1) the ROC curve can be estimated more accurately by both \widehat{R}_E and \widehat{R}_N in the case when the overlap is smaller; and (2) \widehat{R}_N outperforms \widehat{R}_E in all cases when the two populations are normally distributed.

We then try several different types of distributions for both X and Y . More specifically X has one of the following four distributions: (i) $N(0.5, 1)$; (ii) $T/\sqrt{3} + 0.5$, where T is a t -distribution variable with $df = 3$; (iii) $-E + 1.5$, where E is an exponential-distribution variable with mean 1; and (iv) $E - 0.5$. Notice that the four distributions have the same mean 0.5 and the same standard deviation 1. The first distribution is symmetric short-tailed, the second one is symmetric long-tailed, the third one is skewed to the left and the last one is skewed to the right. The distribution of Y is chosen from the first three distributions but with mean 0 (by subtracting 0.5). From Section 1, a person is classified as diseased if that person's measurement value is lower than the cutpoint. Therefore if the diseased population distribution is skewed, it is often skewed to the left, reflecting the fact that a number of people have the disease in quite serious stages. So we only consider a left-skewed distribution for Y . The results are shown in Figure 3.4 to compare \widehat{R}_E with \widehat{R}_N for all twelve combinations of the X and Y distributions. From the plots, we can see that \widehat{R}_N outperforms \widehat{R}_E in almost all cases except several occasions in plots (c), (f) and (g) in which \widehat{R}_E performs a little bit better than \widehat{R}_N when $n=5$ and $m \geq 10$.

As Figure 3.3 indicated, the overlap between the diseased and non-diseased population distributions affects the performance of the ROC curve estimators. Next we increase the mean of the non-diseased population distribution in the above example by 1.0 and keep other properties of the two distributions unchanged (therefore the overlap between the two distributions is decreased).

Then the results corresponding to those in Figure 3.4 are presented in Figure 3.5. It can be seen from the plots that \widehat{R}_N still outperforms \widehat{R}_E when the non-diseased population distribution is $N(1.5, 1)$ or $E + 0.5$ (cf. the first and last columns of Figure 3.5). When the non-diseased population distribution is $T/\sqrt{3} + 1.5$ (which is heavy-tailed) or $-E + 2.5$ (which is skewed to the left), \widehat{R}_N does not perform well comparing with \widehat{R}_E , especially when n (sample size of the sample from the non-diseased population) is small. This could be explained by the fact that the Harrell-Davis quantile estimator $\widehat{F}_{H-D}^{-1}(s)$ (defined in (2.1)), which is used in \widehat{R}_N , does not perform well when s is close to 0 or 1 (cf. Harrell and Davis (1982) for the related discussion).

When the non-diseased population distribution is $T/\sqrt{3} + 1.5$ and $n = 20$, the squared errors (SE) of the Harrell-Davis quantile estimator and the empirical quantile estimator (defined in (1.3)) at 200 equally spaced s positions are displayed in Figure 3.6(a). As before, each of these SE values is an average of 1000 replications. The ratio of the SE value of the Harrell-Davis quantile estimator to the SE value of the empirical quantile estimator at each s position is displayed in Figure 3.6(b). From the plots, it can be seen that the Harrell-Davis quantile estimator does not perform well comparing with the empirical quantile estimator when s is small (say $s < 0.2$) or when s is large (say $s > 0.8$).

When the overlap between the diseased and non-diseased population distributions is small, the estimated ROC curve is based mainly on the part of the quantile estimator with small s value. For this reason, \widehat{R}_N would not perform satisfactorily in the case when the overlap is small. Figure 3.6(c) presents the SE values of both \widehat{R}_N and \widehat{R}_E when the non-diseased population distribution is $T/\sqrt{3} + 1.5$, the diseased population distribution is $T/\sqrt{3}$ and $m = n = 20$. It can be seen from this plot that (1) \widehat{R}_N does not perform well when s is small and (2) the two ROC curve estimators both perform well when s is large. Consequently the averaged SE value of \widehat{R}_N (namely, the ASE value of \widehat{R}_N defined at the beginning of this section) across different s levels is larger than the averaged SE value of \widehat{R}_E as shown by Figure 3.5(f). As a comparison, the corresponding results when the non-diseased population distribution is $T/\sqrt{3} + 0.5$, the diseased population distribution and the sample sizes are kept unchanged are presented in Figure 3.6(d) (please notice that the overlap between the two population distributions in this case is larger than the overlap in the case of Figure 3.6(c)). From the plot, we can see that the SE value of \widehat{R}_N is briefly above the SE value of \widehat{R}_E when s is small and then it is quite consistently under the SE value of \widehat{R}_E . Consequently, the averaged SE value of \widehat{R}_N across different s levels is smaller than the averaged SE value of \widehat{R}_E as shown by Figure

3.4(f). We also performed simulations in the case when the non-diseased population distribution is $-E + 2.5$. The results are similar and therefore are not presented here.

4 Concluding Remarks

We have presented a nonparametric procedure to fit ROC curves. The ROC curve estimator by this procedure is explicitly defined by a composite of a distribution function estimator and a quantile function estimator. It is therefore easy to estimate the sensitivity value for a given specificity value by this procedure. Based on a simulation study, this procedure outperforms the empirical procedure in almost all cases when the overlap between the diseased and non-diseased population distributions is large. When the overlap between the two distributions is small, our procedure performs well when the non-diseased population distribution is symmetric short-tailed or skewed to the right. It does not perform well when the non-diseased population distribution is symmetric heavy-tailed or skewed to the left.

Acknowledgements The authors are grateful to a referee for many helpful comments which greatly improved the article.

REFERENCES

- Altman, N., and Leger, C. (1995). Bandwidth selection for kernel distribution function estimation. *Journal of Statistical Planning and Inference*, **46**, 195-214.
- Cai, Z., and Roussas, G.G. (1998). Efficient estimation of a distribution function under quadrant dependence. *Scandinavian Journal of Statistics*, **25**, 211-224.
- Campbell, G. (1994). General methodology I: advances in statistical methodology for the evaluation of diagnostic and laboratory tests. *Statistics in Medicine*, **13**, 499-508.
- DeLong, E.R., DeLong, D.M., and Clarke-Pearson, D.L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics*, **44**, 837-845.
- Dielman, T., Lowry, C., and Pfaffenberger, R. (1994). A comparison of quantile estimators. *Communications in Statistics - Simulations*, **23**, 355-371.

- Eubank, R.L. (1985). Quantiles. In *Encyclopedia of Statistical Science*, **6** (S. Kotz and N.L. Johnson, eds.), 424-432. Wiley: New York.
- Falk, M. (1983). Relative efficiency and deficiency of kernel type estimators of smooth distribution functions. *Statistica Neerlandica*, **37**, 73-83.
- Hanley, J.A. and McNeil, B.J. (1982). The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology*, **143**, 29-36.
- Harrell, F.E., and Davis, C.E. (1982). A new distribution-free quantile estimator. *Biometrika*, **69**, 635-640.
- Hin, L.Y., Lau, T.K., Rogers, M., and Chang, A.M.Z. (1997). Antepartum and intrapartum prediction of cesarean need: risk scoring in singleton pregnancies. *Obstetrics & Gynecology*, **90**, 183-186.
- Le, C.T. (1997). Evaluation of confounding effects in ROC studies. *Biometrics*, **53**, 998-1007.
- Lloyd, C.J. (1998). Using smoothed receiver operating characteristic curves to summarize and compare diagnostic systems. *Journal of the American Statistical Association*, **93**, 1356-1364.
- Majumder, K.L., and Bhattacharjee, G.P. (1973). The incomplete beta integral (Algorithm AS63). *Applied Statistics*, **22**, 409-411.
- Metz, C.E., and Pan, X. (1999). "Proper" binormal ROC curves: theory and maximum-likelihood estimation. *Journal of Mathematical Psychology*, **43**, 1-33.
- Pan, X., and Metz, C.E. (1997). The "proper" binormal model: parametric ROC curve estimation with degenerate data. *Academic Radiology*, **4**, 380-390.
- Parrish, R.S. (1990). Comparison of quantile estimators in normal sampling. *Biometrics*, **46**, 247-257.
- Reiss, R.D. (1981). Non-parametric estimation of smooth distribution functions. *Scandinavian Journal of Statistics*, **8**, 116-119.
- Zweig, M.H., and Campbell, G. (1993). Receiver-Operative Characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clinical Chemistry*, **39**, 561-577.

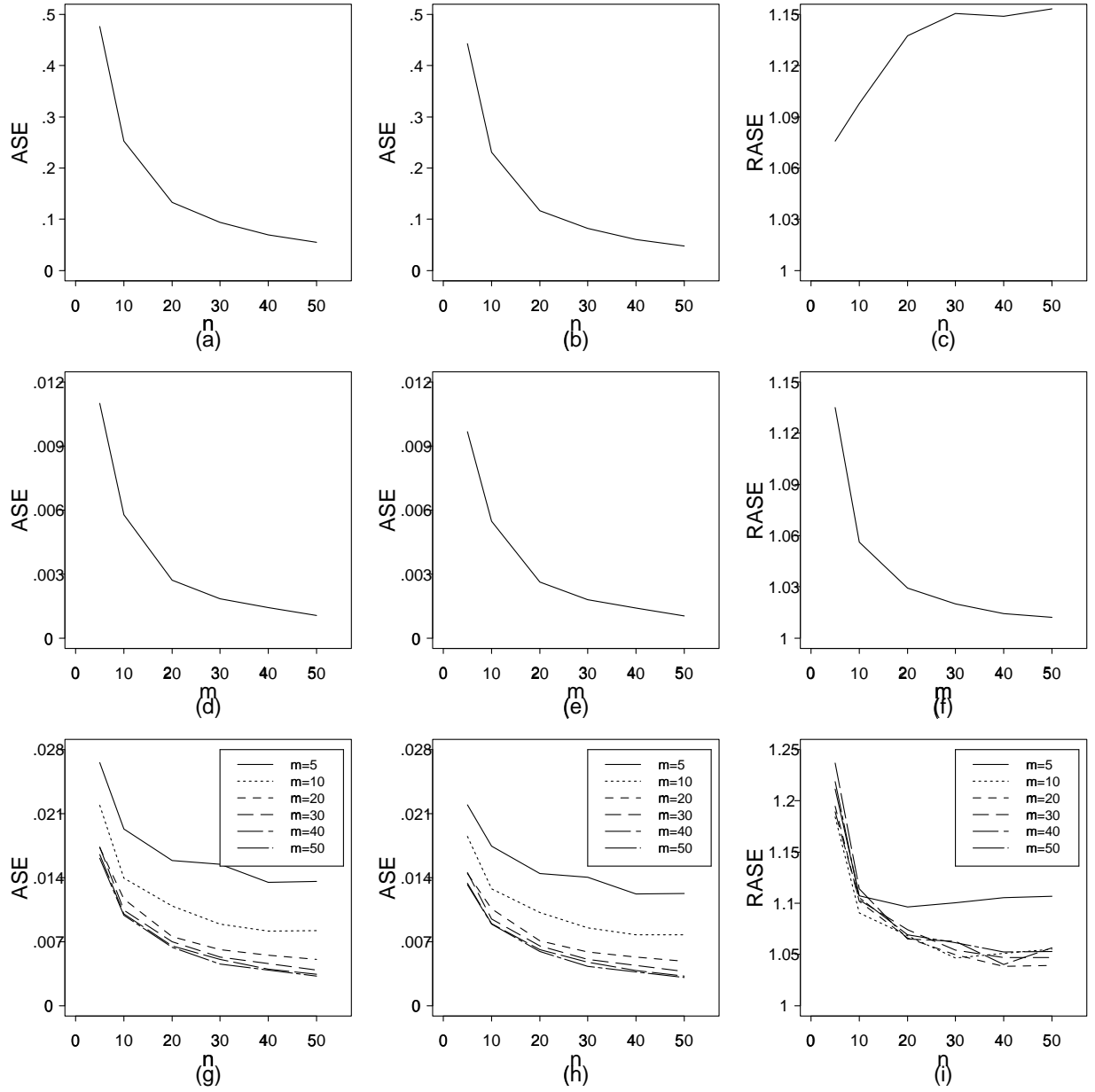


Figure 3.2: (a) $ASE(\hat{F}_n^{-1}, F^{-1})$; (b) $ASE(\hat{F}_{H-D}^{-1}, F^{-1})$; (c) $ASE(\hat{F}_n^{-1}, F^{-1})/ASE(\hat{F}_{H-D}^{-1}, F^{-1})$; (d) $ASE(\hat{G}_m, G)$; (e) $ASE(\hat{G}_{ker}, G)$; (f) $ASE(\hat{G}_m, G)/ASE(\hat{G}_{ker}, G)$; (g) $ASE(\hat{R}_E, R)$; (h) $ASE(\hat{R}_N, R)$; (i) $ASE(\hat{R}_E, R)/ASE(\hat{R}_N, R)$.

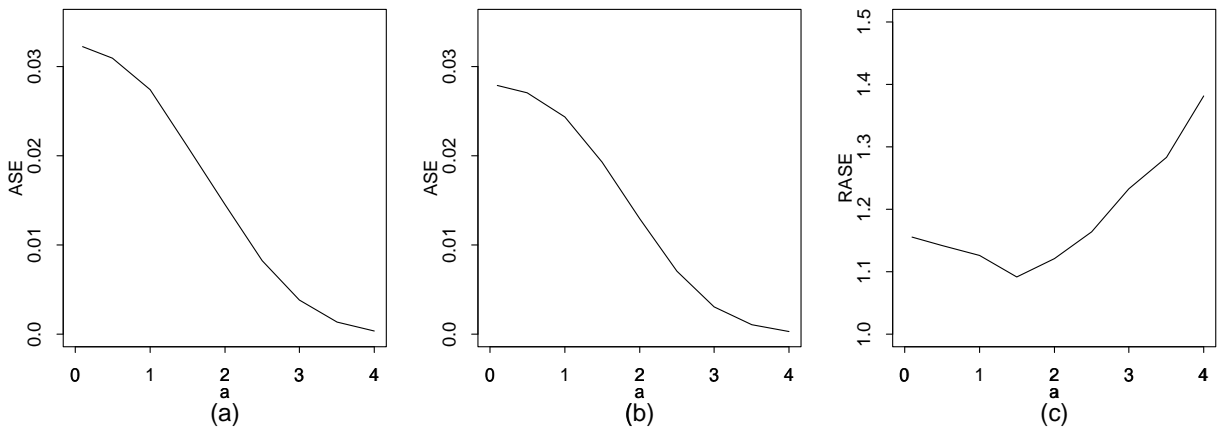


Figure 3.3: (a) $ASE(\hat{R}_E, R)$; (b) $ASE(\hat{R}_N, R)$; (c) $ASE(\hat{R}_E, R)/ASE(\hat{R}_N, R)$.

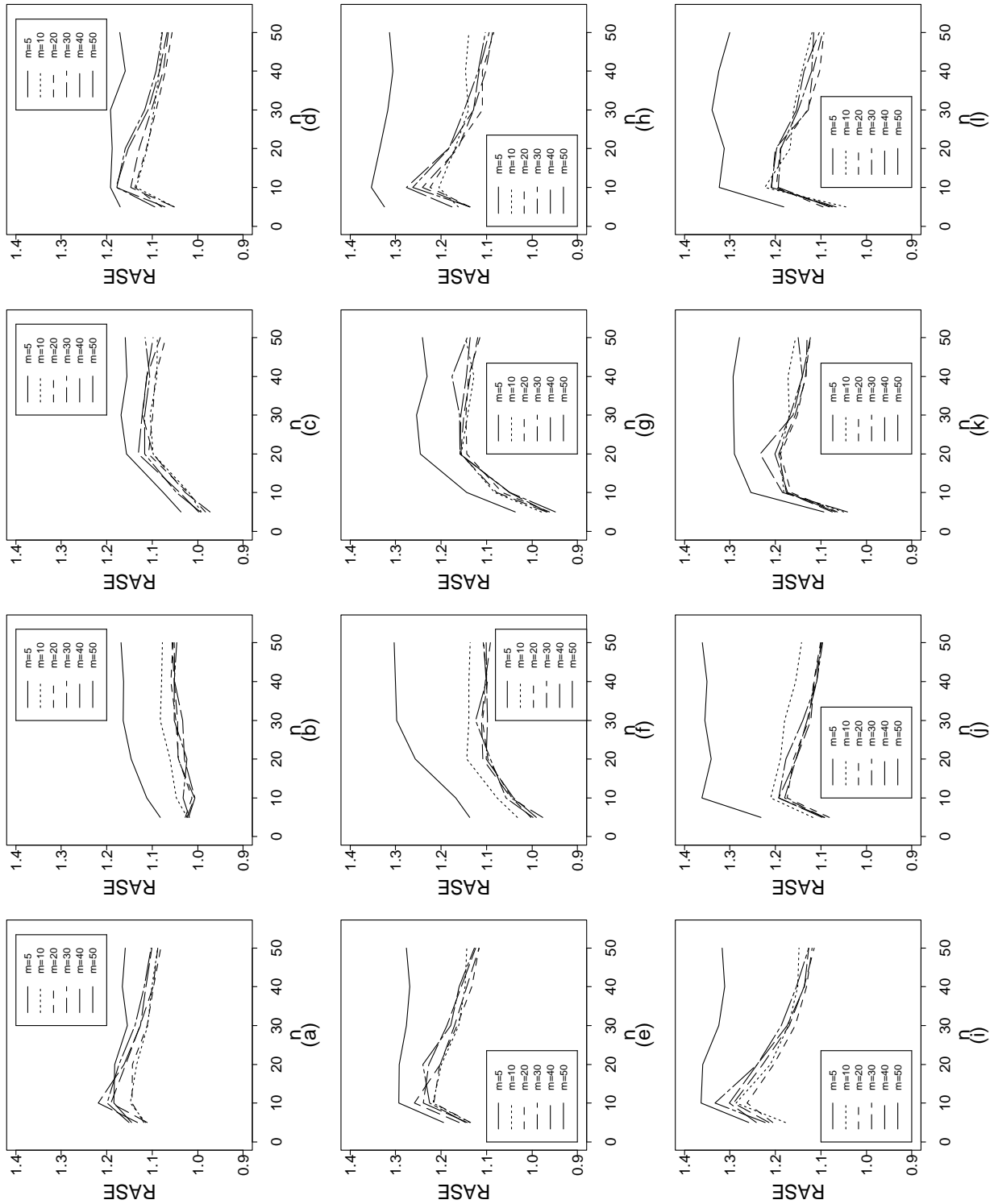


Figure 3.4: $ASE(\hat{R}_E, R)/ASE(\hat{R}_N, R)$. (a) $X \sim N(0.5, 1)$, $Y \sim N(0, 1)$; (b) $X \sim T/\sqrt{3} + 0.5$, $Y \sim N(0, 1)$; (c) $X \sim -E + 1.5$, $Y \sim N(0, 1)$; (d) $X \sim E - 0.5$, $Y \sim N(0, 1)$; (e) $X \sim N(0.5, 1)$, $Y \sim T/\sqrt{3}$; (f) $X \sim T/\sqrt{3} + 0.5$, $Y \sim T/\sqrt{3}$; (g) $X \sim -E + 1.5$, $Y \sim T/\sqrt{3}$; (h) $X \sim E - 0.5$, $Y \sim T/\sqrt{3}$; (i) $X \sim N(0.5, 1)$, $Y \sim -E + 1$; (j) $X \sim T/\sqrt{3} + 0.5$, $Y \sim -E + 1$; (k) $X \sim -E + 1.5$, $Y \sim -E + 1$; (l) $X \sim E - 0.5$, $Y \sim -E + 1$; where T is a t -distribution variable with $df = 3$ and E is an exponential distribution variable with mean 1.

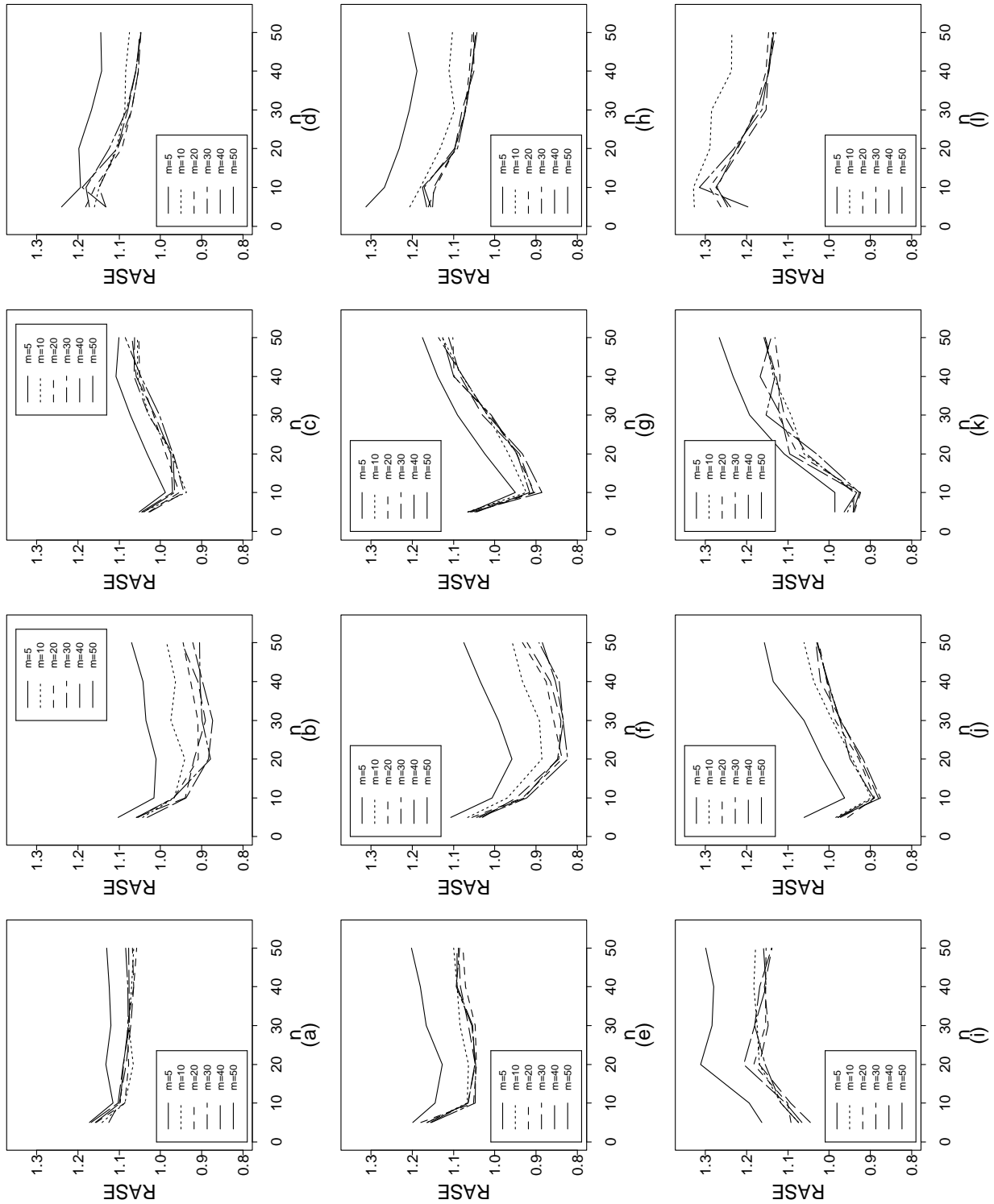


Figure 3.5: $ASE(\hat{R}_E, R)/ASE(\hat{R}_N, R)$. (a) $X \sim N(1.5, 1)$, $Y \sim N(0, 1)$; (b) $X \sim T/\sqrt{3} + 1.5$, $Y \sim N(0, 1)$; (c) $X \sim -E + 2.5$, $Y \sim N(0, 1)$; (d) $X \sim E + 0.5$, $Y \sim N(0, 1)$; (e) $X \sim N(1.5, 1)$, $Y \sim T/\sqrt{3}$; (f) $X \sim T/\sqrt{3} + 1.5$, $Y \sim T/\sqrt{3}$; (g) $X \sim -E + 2.5$, $Y \sim T/\sqrt{3}$; (h) $X \sim E + 0.5$, $Y \sim T/\sqrt{3}$; (i) $X \sim N(1.5, 1)$, $Y \sim -E + 1$; (j) $X \sim T/\sqrt{3} + 1.5$, $Y \sim -E + 1$; (k) $X \sim -E + 2.5$, $Y \sim -E + 1$; (l) $X \sim E + 0.5$, $Y \sim -E + 1$; where T is a t -distribution variable with $df = 3$ and E is an exponential distribution variable with mean 1.

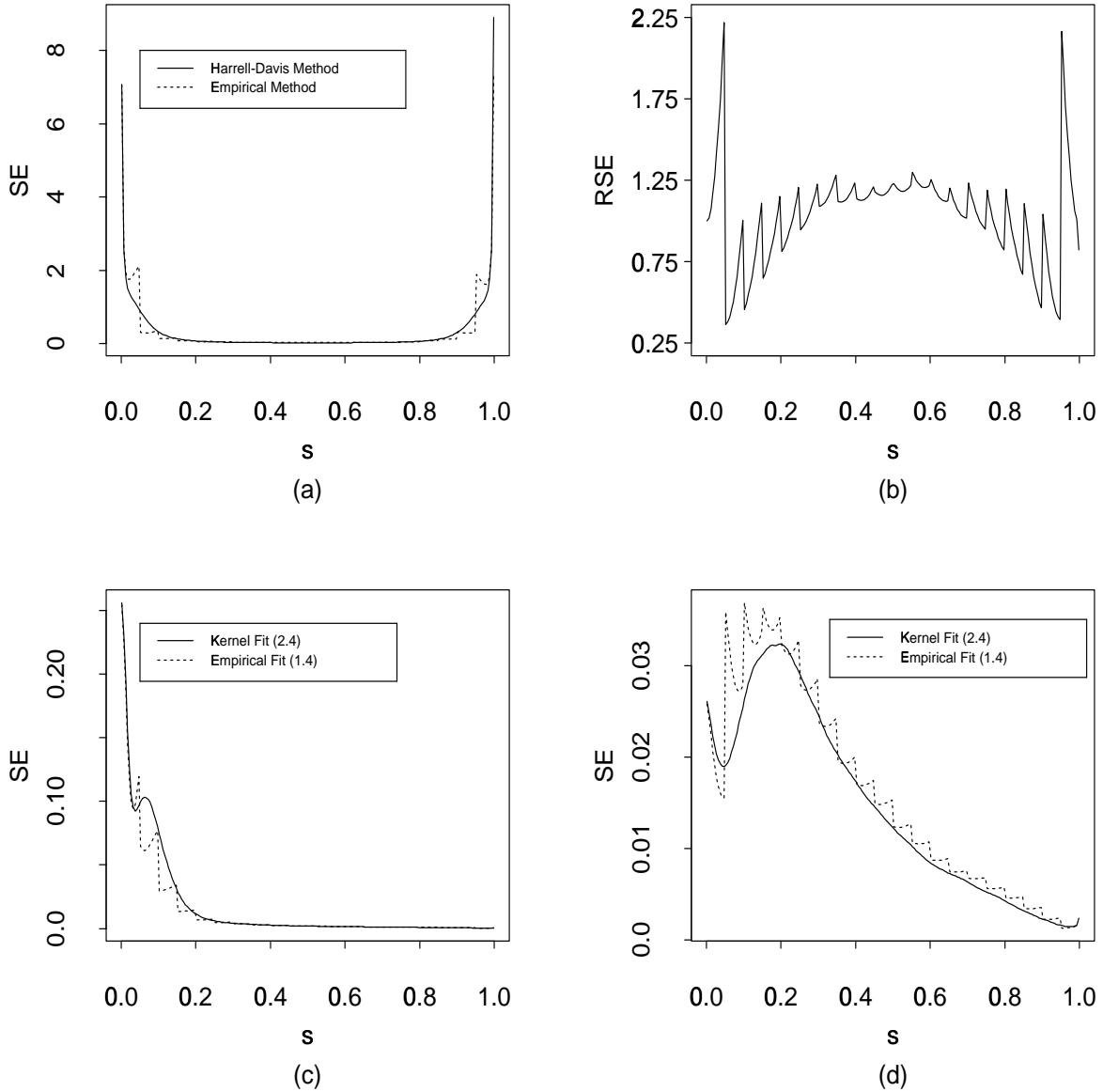


Figure 3.6: (a) The squared errors (SE) of the Harrell-Davis quantile estimator and the empirical quantile estimator when the non-diseased population distribution is $T/\sqrt{3} + 1.5$ and $n = 20$; (b) the ratio of the SE of the Harrell-Davis quantile estimator to the SE of the empirical quantile estimator; (c) the SE values of \hat{R}_N and \hat{R}_E when the non-diseased population distribution is $T/\sqrt{3} + 1.5$; (d) the SE values of \hat{R}_N and \hat{R}_E when the non-diseased population distribution is $T/\sqrt{3} + 0.5$. The diseased population distribution is $T/\sqrt{3}$ and $m = n = 20$ in plots (c) and (d).