Biometrika (2004), **91**, 3, *pp.* 743–750

Printed in Great Britain

Nonparametric confidence intervals for receiver operating characteristic curves

By PETER HALL

Centre for Mathematics and its Applications, Australian National University, Canberra ACT 0200, Australia

peter.hall@maths.anu.edu.au

ROB J. HYNDMAN

Department of Econometrics and Business Statistics, Monash University, Clayton VIC 3800, Australia

rob.hyndman@buseco.monash.edu.au

AND YANAN FAN

Department of Statistics, University of New South Wales, Sydney NSW 2052, Australia

yanan@maths.unsw.edu.au

Summary

We study methods for constructing confidence intervals and confidence bands for estimators of receiver operating characteristics. Particular emphasis is placed on the way in which smoothing should be implemented, when estimating either the characteristic itself or its variance. We show that substantial undersmoothing is necessary if coverage properties are not to be impaired. A theoretical analysis of the problem suggests an empirical, plug-in rule for bandwidth choice, optimising the coverage accuracy of interval estimators. The performance of this approach is explored. Our preferred technique is based on asymptotic approximation, rather than a more sophisticated approach using the bootstrap, since the latter requires a multiplicity of smoothing parameters all of which must be chosen in nonstandard ways. It is shown that the asymptotic method can give very good performance.

Some key words: Bandwidth selection; Binary classification; Kernel estimator; Receiver operating characteristic curve.

1. Introduction

A receiver operating characteristic curve is often used to describe the performance of a diagnostic test which classifies individuals into either group G_1 or group G_2. It is most commonly used with medical data where, for example, G_1 may contain individuals with a disease and G_2 those without the disease.

We assume that the diagnostic test is based on a continuous measurement T and that a person is classified as G_1 if $T \geqslant \tau$ and G_2 otherwise. Let $G(t) = \mathrm{pr}(T \leqslant t | G_1)$ and $F(t) = \mathrm{pr}(T \leqslant t | G_2)$ denote the distribution functions of T for each group. Then the receiver operating characteristic curve is defined as $R(p) = 1 - G\{F^{-1}(1-p)\}$, where $0 \leqslant p \leqslant 1$.

There is a rapidly-growing literature on methods for estimating the plots, including nonparametric techniques based on kernel ideas (Zou et al., 1997; Hall & Hyndman, 2003) and local linear smoothing (Peng & Zhou, 2004). Claeskens et al. (2003) have considered empirical likelihood methods for constructing confidence intervals.

Against the background of this growing interest in both point and interval estimation, the present paper shows how the bandwidth used to construct an estimator of R influences the performance of pointwise confidence bands. For example, we demonstrate that bandwidths which are appropriate for point or curve estimation are an order of magnitude too large for good coverage accuracy. We discuss asymptotic and bootstrap methods for confidence bands, but favour the asymptotic approach, since bootstrap methods require a multiplicity of decisions about smoothing, all of them needing nonstandard solutions. Furthermore, we shall show that the theoretical analysis which leads to our conclusions about undersmoothing can also be used to develop an explicit and effective device for selecting the correct amount of smoothing for confidence bands. The performance of this approach is demonstrated using both numerical simulation and theory.

The methods of this paper cannot easily be extended to simultaneous confidence bands, which are governed by properties more closely related to extreme-value theory than to the normal distribution; see Hall (1991) for discussion of bootstrap approximations to simultaneous bands. Moreover, there appears to be substantially greater interest in pointwise bands than in simultaneous bands, apparently because many people find the former more easily interpreted.

2. Methodology

2·1. *Distribution estimators*

Suppose we are given independent random samples $\mathscr{X}=\{X_1,\ldots,X_m\}$ and $\mathscr{Y}=\{Y_1,\ldots,Y_m\}$ from distributions with respective distribution functions F and G. Let $\hat{F}_{\text{emp}}$ and $\hat{G}_{\text{emp}}$ denote the corresponding empirical distribution functions. For example, $\hat{F}_{\text{emp}}(x)=m^{-1}\sum_i I(X_i\leqslant x)$, where $I(\mathscr{E})$ denotes the indicator function of an event $\mathscr{E}$. We could estimate the function, $R(p)$, by $\hat{R}(p)=1-\hat{G}_{\text{emp}}\{\hat{F}^{-1}_{\text{emp}}(1-p)\}$. However, $\hat{F}_{\text{emp}}$ and $\hat{G}_{\text{emp}}$ are discontinuous, and, especially if the sample sizes m and n differ, $\hat{R}(p)$ can have a very erratic appearance.

For this reason, and another given later in this section, it can be advantageous to smooth $\hat{F}_{\text{emp}}$ and $\hat{G}_{\text{emp}}$ prior to calculating the estimator of $R(p)$. To this end, let L be a known, smooth distribution function, let h_1 and h_2 denote bandwidths, and put

$$\hat{F}(x)=m^{-1}\sum_{i=1}^{m}L\left(\frac{x-X_i}{h_1}\right),\quad \hat{G}(y)=n^{-1}\sum_{i=1}^{n}L\left(\frac{y-Y_i}{h_2}\right). \tag{2·1}$$

Then $\hat{F}$ and $\hat{G}$ are smoothed versions of $\hat{F}_{\text{emp}}$ and $\hat{G}_{\text{emp}}$, respectively. Their derivatives, $\hat{f}=\hat{F}'$ and $\hat{g}=\hat{G}'$, are conventional kernel estimators of the densities $f=F'$ and $g=G'$, computed using the kernel $K=L'$. Optimal choice of bandwidth for $\hat{F}$ and $\hat{G}$ is quite different from that which is appropriate for $\hat{f}$ and $\hat{g}$, and indeed h_1 and h_2 at (2·1) can be often chosen quite small without seriously hindering perfomance; see for example Altman & Léger (1995) and Bowman et al. (1998).

The kernel estimate of $R(p)$ is $\hat{R}(p)=1-\hat{G}\{\hat{F}^{-1}(1-p)\}$. Bandwidth choice for $\hat{R}(p)$ has been considered by Hall & Hyndman (2003).

In §§ 2·2 and 2·3 we suggest asymptotic and bootstrap methods, respectively, for constructing pointwise confidence intervals for $R(p)$. The former are based on the normal distribution, and so make no attempt to capture the skewness of the estimator of $R(p)$. The latter have an opportunity for capturing skewness, but in both cases optimal performance can only be realised if the distribution function estimators are smoothed. To appreciate why, note that, while $\hat{F}_{\text{emp}}(x)$ can be represented by a sum of independent and identically distributed random variables, the component variables $I(X_i\leqslant x)$ are lattice-valued, and in such cases the bootstrap can fail to capture the main effects of skewness (Singh, 1981). Incorporation of a little smoothing, for example using $\hat{F}$ rather than $\hat{F}_{\text{emp}}$, can overcome these difficulties. This has the added advantage of enhancing the appearance of an estimate of $R(p)$.

2·2. *Asymptotic confidence intervals*

It can be shown that, to a first-order approximation, if $0 < t < 1$ then $\hat{G}\{\hat{F}^{-1}(t)\} - G\{F^{-1}(t)\}$ is distributed as

$$\hat{G}\{\hat{F}^{-1}(t)\} - G\{F^{-1}(t)\} \simeq \hat{G}\{F^{-1}(t)\} - G\{F^{-1}(t)\} - \frac{g\{F^{-1}(t)\}}{f\{F^{-1}(t)\}}[\hat{F}\{F^{-1}(t)\} - t]; \qquad (2·2)$$

see for example Hsieh & Turnbull (1996). For the relatively small bandwidths that would be used to construct $\hat{F}$ and $\hat{G}$, the quantity on the right-hand side of (2·2) is asymptotically normally distributed with zero mean and variance given by

$$\sigma(t)^2 = n^{-1} G\{F^{-1}(t)\}[1 - G\{F^{-1}(t)\}] + m^{-1} \frac{g\{F^{-1}(t)\}^2}{f\{F^{-1}(t)\}^2} t(1-t). \qquad (2·3)$$

Here we have made use of the assumption that $\mathscr{X}$ and $\mathscr{Y}$ are independent samples of independent data.

Replacing F, G, f and g at (2·3) by respective estimators $\hat{F}$, $\hat{G}$, $\tilde{f}$ and $\tilde{g}$, we obtain an estimator of σ:

$$\hat{\sigma}(t)^2 = n^{-1} \hat{G}\{\hat{F}^{-1}(t)\}[1 - \hat{G}\{\hat{F}^{-1}(t)\}] + m^{-1} \frac{\tilde{g}\{\hat{F}^{-1}(t)\}^2}{\tilde{f}\{\hat{F}^{-1}(t)\}^2} t(1-t). \qquad (2·4)$$

We might take $\tilde{f}$ and $\tilde{g}$ here to be simply the estimators $\hat{f}$ and $\hat{g}$, noted earlier. However, a substantially different size of bandwidth can be necessary when optimising confidence intervals for coverage accuracy, relative to that which is appropriate when constructing distribution or density estimators with good pointwise accuracy. We recognise this by using 'tilde' rather than 'hat' notation. For future reference, let h_f and h_g denote the bandwidths used for $\tilde{f}$ and $\tilde{g}$:

$$\tilde{f}(x) = \frac{1}{mh_f} \sum_{i=1}^{m} K\left(\frac{x - X_i}{h_f}\right), \quad \tilde{g}(y) = \frac{1}{nh_g} \sum_{i=1}^{n} K\left(\frac{y - Y_i}{h_g}\right). \qquad (2·5)$$

One-sided, asymptotic, $(1-\alpha)$-level confidence intervals for $R(p)$ are therefore given by $(\hat{R}(p) - z_\alpha \hat{\sigma}(1-p), 1)$ and $(0, \hat{R}(p) + z_\alpha \hat{\sigma}(1-p))$, where $z_\alpha > 0$ is the upper $1-\alpha$ point of the standard normal distribution. A two-sided confidence interval has of course endpoints $\hat{R}(p) \pm z_{\alpha/2} \hat{\sigma}(1-p)$. Here, $\hat{R}(p)$ is based on (2·1) and (2·2), but it does not necessarily use the same bandwidths as are used in (2·4). In our numerical examples, we estimate $\hat{R}(p)$ using the bandwidth proposal of Hall & Hyndman (2003).

2·3. *Bootstrap confidence intervals*

An alternative approach to constructing interval estimators is to approximate the distribution of

$$S = [G\{F^{-1}(t)\} - \hat{G}\{\hat{F}^{-1}(t)\}]/\hat{\sigma},$$

using the bootstrap and Monte Carlo simulation. To be specific, draw data $\mathscr{X}^* = \{X_1^*, \ldots, X_m^*\}$ and $\mathscr{Y}^* = \{Y_1^*, \ldots, Y_n^*\}$ randomly, without replacement, from distributions with respective densities $\check{f}$ and $\check{g}$, where $\check{f}$ and $\check{g}$ are smoothed estimators of f and g and are computed from $\mathscr{X}$ and $\mathscr{Y}$, respectively. Compute the bootstrap versions, $\hat{F}^*$, $\hat{G}^*$, $\tilde{f}^*$ and $\tilde{g}^*$ say, of $\hat{F}$, $\hat{G}$, $\tilde{f}$ and $\tilde{g}$; let $\hat{\sigma}^*$ denote the version of $\hat{\sigma}$ at (2·4) that is obtained on replacing the latter estimators by their bootstrap forms; write $\check{F}$ and $\check{G}$ for the respective distribution functions corresponding to the densities $\check{f}$ and $\check{g}$; and put

$$S^* = [\check{G}\{\check{F}^{-1}(t)\} - \hat{G}^*\{(\hat{F}^*)^{-1}(t)\}]/\hat{\sigma}^*. \qquad (2·6)$$

Then the distribution of S^*, conditional on the original data $\mathscr{Z} = (\mathscr{X}, \mathscr{Y})$, is an approximation to the unconditional distribution of S.

In particular, we may compute $\hat{z}_\alpha = \hat{z}_\alpha(\mathscr{X})$ as the solution of the equation $\mathrm{pr}(S^* \leqslant \hat{z}_\alpha | \mathscr{X}) = \alpha$, for $0 < \alpha < 1$, and take one-sided, $(1-\alpha)$-level confidence intervals for $R(p)$ to be $(\hat{R}(p) - \hat{z}_\alpha \hat{\sigma}(1-p), 1)$ and $(0, \hat{R}(p) - \hat{z}_{1-\alpha}\hat{\sigma}(1-p))$. These are of course percentile-t intervals.

We have introduced additional density estimators, $\check{f}$ and $\check{g}$, rather than use the existing estimators, since it is initially far from clear what the appropriate levels of smoothing in the bootstrap resampling step should be.

3. Coverage probabilities

3·1. *Effect of bandwidth choice on asymptotic intervals*

Let $\alpha \in (\frac{1}{2}, 1)$, and define z_α by $\Phi(z_\alpha) = 1 - \alpha$, where Φ denotes the standard normal distribution function. Also, let $y_1 = \hat{F}^{-1}(t)$ and $y = F^{-1}(t)$. Since $\{\hat{G}(y_1) - G(y)\}/\hat{\sigma}$ is asymptotically $N(0, 1)$ then examples of asymptotic confidence intervals for $R(p) = 1 - G\{F^{-1}(1-p)\}$ are given by

$$(-\infty, \hat{R}(p) + \hat{\sigma} z_\alpha], \quad [\hat{R}(p) - \hat{\sigma} z_\alpha, -\infty), \quad [\hat{R}(p) - \hat{\sigma} z_{\alpha/2}, \hat{R}(p) + \hat{\sigma} z_{\alpha/2}]. \tag{3·1}$$

The coverage probability of each converges to α as $n \to \infty$.

In familiar semiparametric problems, such as confidence intervals for a population mean, the three intervals at (3·1) would have coverage errors of sizes $n^{-1/2}$, $n^{-1/2}$ and n^{-1}, respectively. In particular, the terms in $n^{-1/2}$ that dominate coverage-error formulae for one-sided intervals cancel, in the two-sided case, through a fortuitous parity property, and then second-order terms, of size n^{-1}, prevail. In the present setting, however, such a simple account of coverage accuracy is prevented because $\hat{\sigma}$ involves a nonparametric component, depending critically on the bandwidths h_f and h_g used to construct $\tilde{f}$ and $\tilde{g}$ at (2·5), and employed to compute $\hat{\sigma}$. It can be shown that, if h_f and h_g are chosen to be of conventional size, $n^{-1/5}$, appropriate for point estimation of f and g, then the coverage error of each of the confidence intervals at (3·1) is of size $n^{-2/5}$, which falls short even of the level $n^{-1/2}$ that is available in the one-sided case in a classical setting.

That this is true even for the third, two-sided interval at (3·1) follows because the leading terms which introduce h_f and h_g to coverage-error formulae do not enjoy the classical parity property. As a result, errors of size $n^{-2/5}$ persist for each of the three intervals at (3·1). They compound, rather than cancel, in passing from one-sided to two-sided intervals. There are, or course, two other bandwidths, h_1 and h_2, used to construct $\hat{F}$ and $\hat{G}$ at (2·1). These, however, have only a minor impact, and can be chosen within a wide range without seriously affecting coverage error.

These results motivate a careful analysis of the impact that choosing h_f and h_g has on coverage accuracy. We shall show that it is optimal to select these bandwidths to be constant multiples of $m^{-1/3}$ and $n^{-1/3}$, respectively, and we shall suggest formulae for the constants. With this choice of the bandwidths, the coverage errors of the one-sided intervals at (3·1) are of size $n^{-1/2}$, reducing to $n^{-2/3}$ in the two-sided setting. Thus, accuracy in the one-sided case coincides with that in classical problems, while in the two-sided setting it is a little less than in the classical case, but still better than for one-sided intervals.

Next we describe our main theoretical results. Put $\rho = n/m$, $\kappa = \int K(u)^2 du$, $\kappa_2 = \int u^2 K(u) du$ and

$$a = \frac{\rho\{g(y)/f(y)\}^2 t(1-t)}{G(y)\{1 - G(y)\} + \rho\{g(y)/f(y)\}^2 t(1-t)}. \tag{3·2}$$

Note that $0 < a < 1$. Define the even, quadratic polynomials

$$q_f(x) = \kappa(3 - a - ax^2) + 2K(0)(a + ax^2 - 1), \quad q_g(x) = \kappa(a - 1 - ax^2).$$

For $\psi = f$ or g, put $p_\psi(x) = ax\{q_\psi(x) - \kappa_2 m h_\psi^3 \psi''(y)\}/2\psi(y)$, an odd, cubic polynomial. Construct $\tilde{f}$ and $\tilde{g}$ using the kernel K and the respective bandwidths h_f and h_g. Then, provided h_f and h_g are of respective sizes $m^{-1/3}$ and $n^{-1/3}$,

$$\mathrm{pr}[\{\hat{G}(y_1) - G(y)\}/\hat{\sigma} \leqslant x] = \Phi(x) + n^{-1/2} p(x)\phi(x) + \frac{p_f(x)}{2mh_f}\phi(x) - \frac{p_g(x)}{2nh_g}\phi(x) + o(n^{-2/3}), \tag{3·3}$$

where p denotes an even, quadratic polynomial, the coefficients of which do not depend on h_f or h_g, and which involve m and n only through the ratio ρ, remaining bounded as long as ρ is bounded away from zero and infinity. Regularity conditions for (3·3) will be given later in this section, and details of the derivation of (3·3) are available from the authors.

The implications of (3·3) are tied to parity properties of the polynomials p, p_f and p_g. Note that p is even, whereas p_f and p_g are odd, and so (3·3) implies that the two-sided confidence interval, $\mathscr{I} = [\hat{R}(p) - \hat{\sigma} z_{\alpha/2}, \hat{R}(p) + \hat{\sigma} z_{\alpha/2}]$, has coverage probability

$$\mathrm{pr}\,\{R(p) \in \mathscr{I}\} = \alpha + \left\{ \frac{p_f(z_{\alpha/2})}{m h_f} - \frac{p_g(z_{\alpha/2})}{n h_g} \right\} \phi(z_{\alpha/2}) + o(n^{-2/3}). \tag{3·4}$$

Depending on the values of a, α, m, n, $f^{(j)}(y)$ and $g^{(j)}(y)$, for $j = 1, 2$, it can be possible to choose h_f and h_g at (3·4) so that the quantity within braces vanishes. This is not always feasible, however, and a simpler approach is to select h_f and h_g separately, to minimise absolute values of the respective terms within braces. Either approach produces bandwidths of sizes $m^{-1/3}$ and $n^{-1/3}$, respectively; the second approach results in the formulae $h_f = c_f m^{-1/3}$ and $h_g = c_g n^{-1/3}$, where

$$c_f = \theta \left| \frac{q_f(z_{\alpha/2})}{\kappa_2 f''(y)} \right|^{1/3}, \quad c_g = \theta \left| \frac{q_g(z_{\alpha/2})}{\kappa_2 g''(y)} \right|^{1/3}, \tag{3·5}$$

and $\theta = 1$ or $2^{-1/3}$ according as the ratio of the term within modulus signs is positive or negative. This approach to bandwidth choice is also appropriate when constructing the one-sided interval $\mathscr{J} = (-\infty, \hat{R}(p) + \hat{\sigma} z_\alpha]$. There the formulae at (3·5) remain valid, except that $z_{\alpha/2}$ should be replaced by z_α. For a constant bandwidth over the curve, we integrate the numerator and denominator of (3·5) over y.

By way of regularity conditions for (3·3) we require the following: (a) f and g have two continuous derivatives in a neighbourhood of y; (b) neither $f(y)$ nor $g(y)$ vanishes; (c) K is a continuous, symmetric, compactly supported density; (d) the bandwidths h_1 and h_2 used to construct $\hat{F}$ and $\hat{G}$, at (2·1), satisfy $h_j = o(n^{-7/12})$ and $nh_j/\log n \to \infty$ as $n \to \infty$; and (e) the sample-size ratio, ρ, is bounded away from zero and infinity as $n \to \infty$.

The regularity conditions (a)–(e) are mild, and it is clear that except possibly for (d) they are usually assured in practical settings. Moreover, (d) is guaranteed, in most cases of interest, if we choose h_1 and h_2 to be as small as possible subject to the jump discontinuities of $\hat{F}_{\text{emp}}$ and $\hat{G}_{\text{emp}}$ being 'smoothed away' by $\hat{F}$ and $\hat{G}$, respectively, except in the extreme tails. This follows because, away from the tails, the maximum spacing of order statistics is of size $n^{-1} \log n$, and, across the entire distribution, is an order of magnitude larger provided that at least one tail of each of f and g descends to zero. Choosing a bandwidth that is just sufficiently large to smooth away jumps is the approach that is often followed in practice when using kernel methods to estimate a distribution function.

To implement the asymptotic intervals requires ten different smoothing parameters: very small bandwidths h_1 and h_2 for $\hat{F}$ and $\hat{G}$, at (2·1); bandwidths h_f and h_g for $\tilde{f}$ and $\tilde{g}$, at (2·5); bandwidths H_1 and H_2 for estimating $f''(y)$ and $g''(y)$ in (3·5); bandwidths H_f and H_g for estimating $f(y)$ and $g(y)$ in (3·2); and bandwidths H_F and H_G for estimating $F(y)$ and $G(y)$ in (3·2).

In our numerical examples we choose h_1 and h_2 to be 0·25 times the plug-in bandwidths for conditional distribution estimation (Lloyd & Yong, 1999); we choose h_f and h_g using (3·5); H_1 and H_2 are chosen to be optimal assuming f and g are normal, so that $H_1 = (\frac{4}{7})^{1/9} m^{-1/9} s_x$, where s_x is the standard deviation of the $\mathscr{X}$, and H_2 is chosen analogously; we choose H_f and H_g using the Sheather & Jones (1991) plug-in rule; and we choose an analogous plug-in rule for H_F and H_G. R code to carry out these calculations is available from Rob Hyndman.

3·2. *Bootstrap intervals*

A bootstrap version of (3·3) is readily developed. It has the form

$$\mathrm{pr}\,(S^* \leqslant x|\mathscr{X}) = \Phi(x) + n^{-1/2}p(x)\phi(x) + \frac{p_f(x)}{2mh_f}\phi(x) - \frac{p_g(x)}{2nh_g}\phi(x) + o_p(n^{-2/3}), \quad (3{\cdot}6)$$

where S^* is as defined at (2·6). Recall that $\mathscr{X}$ denotes the set of all data X_i and Y_j. The right-hand side of (3·6) is identical to its counterpart at (3·2), except that the remainder is now stochastic.

Result (3·6) follows from a close analogue of (3·3), in which the quantities f, g and their derivatives, appearing in formulae for p_f and p_g, are replaced by their counterparts involving $\check{f}$ and $\check{g}$. In order for (3·6) to follow from this particular expansion it is necessary that $\check{f}$ and $\check{g}$ involve sufficient smoothing to ensure that their second derivatives consistently estimate the second derivatives of f and g, respectively. In mathematical terms this means that the bandwidths used to construct $\check{f}$ and $\tilde{g}$ should converge to zero more slowly than $m^{-1/5}$ and $n^{-1/5}$, respectively. Ideally, in the case of sufficiently smooth densities, the bandwidths should be of sizes $m^{-1/9}$ and $n^{-1/9}$. Thus, oversmoothing is required at this level; conventional bandwidth choices are of sizes $m^{-1/5}$ and $n^{-1/5}$. Without oversmoothing, the bootstrap method described in § 2·3 may not lead to improvements over the asymptotic approach. If sufficient oversmoothing is used, however, then it can be deduced from (3·3) and (3·6) that the bootstrap will produce one- and two-sided confidence intervals with coverage error equal to $o(n^{-2/3})$.

Therefore, choice of bandwidth for constructing the smoothed distribution estimators, $\tilde{F}$ and $\tilde{G}$, from which bootstrap sampling is done is a critical matter. For proper implementation the bootstrap technique requires six quite different, and all nonstandard, smoothing parameters: very small bandwidths h_1 and h_2 for $\hat{F}$ and $\hat{G}$, at (2·1); larger, but still smaller than usual, bandwidths h_f and h_g for $\tilde{f}$ and $\tilde{g}$, at (2·5); and quite large bandwidths for $\check{F}$ and $\check{G}$. This complexity makes the bootstrap approach particularly challenging, and relatively unattractive, to implement.

4. Examples

We computed the actual coverage probability of our confidence intervals using simulations on the following six examples having a range of density shapes.

Example 1: $F = N(0, 1)$; $G = N(1, 1)$.

Example 2: $F = N(0, 1)$; $G = N(2, 2)$.

Example 3: $F = \beta\ (2, 3)$; $G = \beta(2, 4)$.

Example 4: $F = \beta(1{\cdot}2, 3)$; $G = \beta(1{\cdot}2, 2)$.

Example 5: $F = \gamma(2)$; $G = \gamma(3)$.

Example 6: $F = t(5)$; $G = 0{\cdot}2\{t(5) - 1\} + 0{\cdot}8\{t(5) + 1\}$.

Here $\beta(a, b)$ denotes the Beta distribution with density

$$f(x) = \Gamma(a + b)\{\Gamma(a)\Gamma(b)\}^{-1}x^{(a-1)}(1 - x)^{(b-1)} \quad (0 \leqslant x \leqslant 1);$$

$\gamma(a)$ denotes the Gamma distribution with density $x^{a-1}e^{-x}/\Gamma(a)$, for $x > 0$; and $t(v)$ denotes the t distribution with v degrees of freedom.

For each example, we generated 1000 sets of data from F and G, each of sizes $m = n = 100$. Then the curve $\hat{R}(p)$ was computed with bandwidths chosen using the method of Hall & Hyndman (2003). Confidence intervals around the curve were computed using the method outlined in § 2·2.

The proportion of times the confidence interval contained the true $R(p)$ for each p is plotted in Fig. 1. Except in the extreme tails of the distributions, our approach is usually conservative.

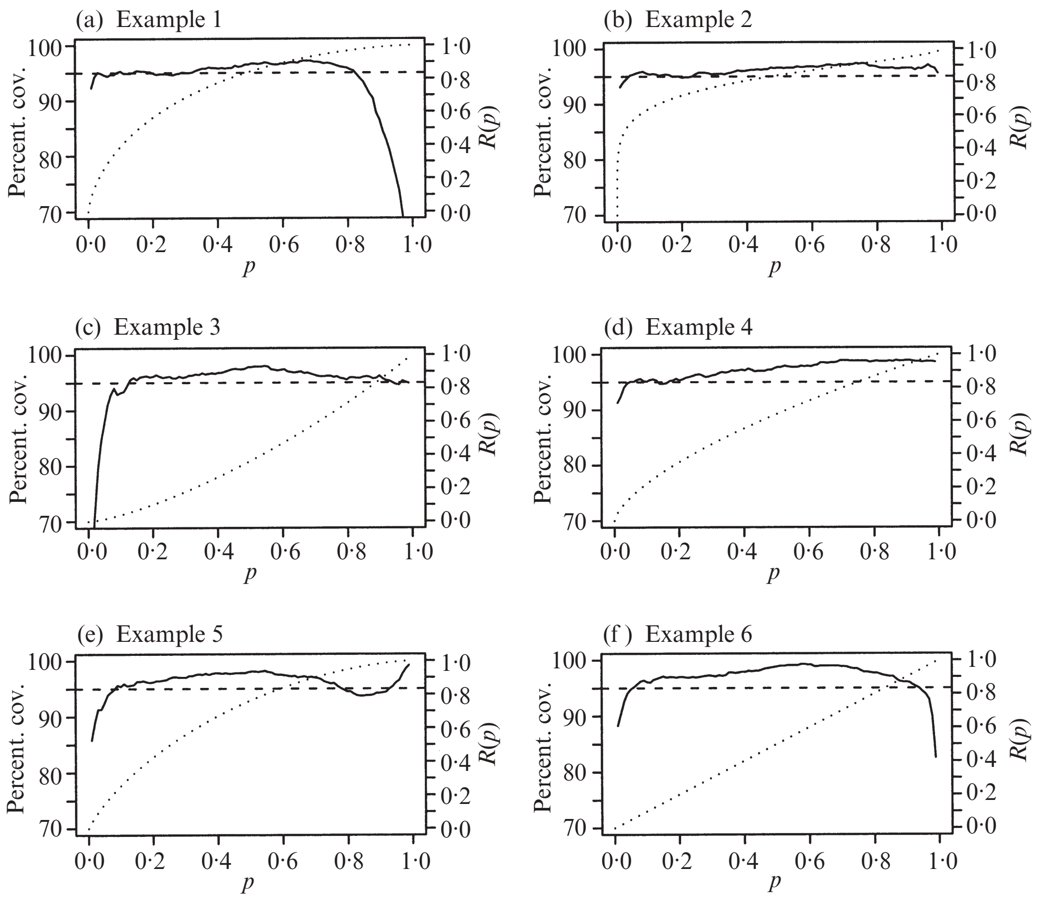

Fig. 1: Examples. Solid line shows actual percentage coverage of asymptotic confidence intervals computed as described in § 2·2. In each example, the percentage is computed from 1000 simulated sets of data. Sample sizes were $m = n = 100$. Nominal coverage was 95%. The dotted line shows the receiver operating characteristic curve in each case, with vertical scale given on the right-hand axis.

Acknowledgement

We thank an anonymous referee for several helpful suggestions.

References

Altman, N. & Léger, C. (1995). Bandwidth selection for kernel distribution function estimation. *J. Statist. Plan. Infer.* **46**, 195–214.

Bowman, A. W., Hall, P. & Prvan, T. (1998). Bandwidth selection for the smoothing of distribution functions. *Biometrika* **85**, 799–808.

Claeskens, G., Jing, B.-Y., Peng, L. & Zhou, W. (2003). Empirical likelihood confidence regions for comparison distributions with ROC curves. *Can. J. Statist.* **31**, 173–90.

Hall, P. (1991). On convergence rates of suprema. *Prob. Theory Rel. Fields* **89**, 447–55.

Hall, P. & Hyndman, R. J. (2003). An improved method for bandwidth selection when estimating ROC curves. *Statist. Prob. Lett.* **64**, 181–9.

Hsieh, F. S. & Turnbull, B. W. (1996). Nonparametric and semiparametric estimation of the receiver operating characteristic curve. *Ann. Statist.* **24**, 25–40.

Lloyd, C. J. & Yong, Z. (1999). Kernel estimators of the ROC curve are better than empirical. *Statist. Prob. Lett.* **44**, 221–8.

PENG, L. & ZHOU, X.-H. (2004). Local linear smoothing of receiver operating characteristic (ROC) curves. *J. Statist. Plan. Infer.* **118**, 129–43.

SHEATHER, S. J. & JONES, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *J. R. Statist. Soc.* B **53**, 683–90.

SINGH, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *Ann. Statist.* **9**, 1187–95.

ZHU, K. H., HALL, W. J. & SHAPIRO, D. E. (1997). Smooth non-parametric receiver operating characteristic (ROC) curves for continuous diagnostic tests. *Statist. Med.* **16**, 2143–56.

[*Received July* 2003. *Revised January* 2004]