Statistica Sinica: Supplement

On Feature Ensemble Optimizing the Sensitivity and Partial ROC Curve

Zheng Zhang

Peking University, P.R.China

Ying Lu and Lu Tian

Stanford University, U.S.A.

Supplementary Material

Appendix A. The Justification for the Equivalence Between $\hat{\beta}_{\hat{w}}$ and $\hat{\beta}_{opt}$.

Firstly, we will show that

$$Q_N(w) = N_0^{-1} \sum_{i=1}^{N} (1 - Y_i) I(\hat{\beta}_w' Z_i \leq \hat{d}_w)$$

is monotone increasing, so that $\hat{w}$ is well defined as the zero-crossing of

$Q_N(w)$. To this end, let $0 < w_1 < w_2$ and we have

$$
\begin{aligned}
& N_1 P_N(w_2) + w_2 N_0 Q_N(w_2) \\
\geq & N_1 P_N(w_1) + w_2 N_0 Q_N(w_1) \\
= & N_1 P_N(w_1) + w_1 N_0 Q_N(w_1) + (w_2 - w_1) N_0 Q_N(w_1) \\
\geq & N_1 P_N(w_2) + w_1 N_0 Q_N(w_2) + (w_2 - w_1) N_0 Q_N(w_1) \\
\Longrightarrow & Q_N(w_2) \geq Q_N(w_1),
\end{aligned}
$$

where

$$P_N(w) = N_1^{-1} \sum_{i=1}^{N} Y_i I(\hat{\beta}_w' Z_i \geq \hat{d}_w).$$

Coupled with the fact that $Q_N(0) = 0$ and $Q_N(+\infty) = 1$, the monotonicity of $Q_N(\cdot)$ suggests that $\hat{w}$ is well defined and $Q_N(\hat{w}) = O_p(N^{-1})$, when there is no tie in $\beta_w' Z_i, i = 1, \cdots, N$.

To show the equivalence between $\hat{\beta}_{\hat{w}}$ and $\hat{\beta}_{opt}$, we note that

$$
\begin{aligned}
& N_0^{-1} \sum_{i=1}^{N} Y_i I(\hat{\beta}_{opt}' Z_i \geq \hat{d}_{opt}) \geq \frac{N_1}{N_0} P_N(\hat{w}) \\
= & N_0^{-1} \left\{ N_1 P_N(\hat{w}) + \hat{w} N_0 Q_N(\hat{w}) \right\} - \hat{w}\pi_0 + O_p(N^{-1}) \\
\geq & N_0^{-1} \left\{ \sum_{i=1}^{N} Y_i I(\hat{\beta}_{opt}' Z_i \geq \hat{d}_{opt}) + \hat{w} \sum_{i=1}^{N} (1 - Y_i) I(\hat{\beta}_{opt}' Z_i < \hat{d}_{opt}) \right\} - \hat{w}\pi_0 + O_p(N^{-1}) \\
= & N_0^{-1} \sum_{i=1}^{N} Y_i I(\hat{\beta}_{opt}' Z_i \geq \hat{d}_{opt}) + O_p(N^{-1}).
\end{aligned}
$$

Thus

$$\left| N_1^{-1} \sum_{i=1}^{N} Y_i I(\hat{\beta}'_{opt} Z_i \geq \hat{d}_{opt}) - P_N(\hat{w}) \right| = O_p(N^{-1}),$$

i.e., the sensitivity $P_N(\hat{w})$ is asymptotically equivalent to that based on $\hat{\beta}_{opt}$.

Appendix B. Asymptotical Properties of $\hat{\beta}_S$ and $\hat{d}_S$

In this Appendix, $g(x) = \log\{1 + \exp(-x)\}$, $\dot{g}(x) = -\{1 + \exp(x)\}^{-1}$ and $\ddot{g}(x) = \exp(x)\{1 + \exp(x)\}^{-2}$. Firstly, we assume the following regularity conditions:

1. $Z_i|Y_i = 0$ and $Z_i|Y_i = 1$ are random vector with a bounded support, which has at least one continuous component and does not belong to any $p - 1$ dimensional hyperplane. Here p is the dimension of Z_i.

2. For any $w > 0$,

$$\begin{aligned} & E\left[\{Y_i + w(1 - Y_i)\} g\{(2Y_i - 1)(\beta' Z_i - d)\}\right] \\ =& E\left\{g(\beta' Z_i - d) \mid Y_i = 1\right\}(1 - \rho) + wE\left\{g(-\beta' Z_i + d) \mid Y_i = 0\right\}\rho \end{aligned}$$

   has an unique minimizer $\beta_0(w)$ and $d_0(w)$. The existence of a finite minimizer can be guaranteed by the assumption that there is no hyperplane in the feature space separating cases from controls almost surely. The uniqueness of the minimizer is true, because the second derivative of the convex objective function

$$E\left(\tilde{Z}_i^{\otimes 2}\ddot{g}\{\beta_0(w)'Z_i - d_0(w)\}\big|Y_i=1\right)(1-\rho)+wE\left(\tilde{Z}_i^{\otimes 2}\ddot{g}\{-\beta_0(w)'Z_i + d_0\}\big|Y_i=0\right)\rho$$

is positive definite, when condition (1) holds, where $\tilde{Z}_i = (-1, Z_i')'$ and $a^{\otimes 2} = aa'$.

3.

$$c_0(w) = E\left\{\frac{1}{1+\exp\{\beta_0(w)'Z_i - d_0(w)\}}\;\middle|\; Y_i = 0\right\}$$

is monotone increasing in w. This condition ensures the existence of the unique root to the estimating equation

$$c_0(w) = \pi_0.$$

It is not difficult to show that

$$E\left[-g\left\{-\beta_0(w)'Z_i + d_0(w)\right\}\;\middle|\; Y_i = 0\right]$$

is increasing in w. To this end, note that for $w_1 \geq w_2$, we have

$$
\begin{aligned}
&(1-\rho)E\left[g\{\beta_0'(w_1) - d_0(w_1)\} \mid Y=1\right] + w_1\rho E\left[g\{-\beta_0(w_1)'Z + d_0(w_1)\} \mid Y=0\right]\\
\leq&(1-\rho)E\left[g\{\beta_0'(w_2) - d_0(w_2)\} \mid Y=1\right] + w_1\rho E\left[g\{-\beta_0(w_2)'Z + d_0(w_2)\} \mid Y=0\right]\\
=&(1-\rho)E\left[g\{\beta_0'(w_2) - d_0(w_2)\} \mid Y=1\right] + \{w_2 + (w_1-w_2)\}\rho E\left[g\{-\beta_0(w_2)'Z + d_0(w_2)\} \mid Y=0\right]\\
\leq&(1-\rho)E\left[g\{\beta_0'(w_1) - d_0(w_1)\} \mid Y=1\right] + w_2\rho E\left[g\{-\beta_0(w_1)'Z + d_0(w_1)\} \mid Y=0\right]\\
&+(w_1-w_2)\rho E\left[g\{-\beta_0(w_2)'Z + d_0(w_2)\} \mid Y=0\right]\\
\Rightarrow&E\left[-g\{-\beta_0(w_1)'Z + d_0(w_1)\}|Y=0\right] \geq E\left[-g\{-\beta_0(w_2)'Z + d_0(w_2)\}|Y=0\right],
\end{aligned}
$$

where $\lim_{N\to\infty} N_0/N = \rho$. Therefore, it is not an unreasonable assumption that

$$c_0(w) = E\left(\exp\left(-\left[g\left\{-\beta_0(w)'Z_i + d_0(w)\right\}\right]\right) \,\middle|\, Y_i = 0\right)$$

is monotone increasing in w as well.

Under the regularity conditions 1-3, there is an unique solution $(\beta_S', \tilde{d}_S, d_S, w_S)'$ to the estimating equation

$$s_0(\beta, \tilde{d}, d, w) = \begin{pmatrix} E\left[Z_i\{Y_i + w(Y_i - 1)\}\dot{g}\{(2Y_i - 1)(\beta' Z_i - \tilde{d})\}\right] \\ E\left[\{Y_i + w(Y_i - 1)\}\dot{g}\{(2Y_i - 1)(\beta' Z_i - \tilde{d})\}\right] \\ E\left[\{1 + \exp(-\tilde{d} + \beta' Z_i)\}^{-1} \mid Y_i = 0\right] - \pi_0 \\ P(\beta' Z_i - d < 0 | Y_i = 0) - \pi_0 \end{pmatrix} = 0. \qquad \text{(S0.1)}$$

We further assume that

4 $(\beta_S', \tilde{d}_S, d_S, w_S)'$ is an interior point of the parameter space $\Omega = \Omega_\beta \times \Omega_d \times \Omega_d \times \Omega_\pi$, which is a compact set of $R^{p+2} \times R^+$, where $R^+ = (0, \infty)$.

5 Let A_0 be the first order derivative of $s_0(\beta, \tilde{d}, d, w)$ at $(\beta_S', \tilde{d}_S, d_S, w_S)'$. A_0 is a non-singular matrix.

Let

$$S_N(\beta, \tilde{d}, d, w) = N^{-1} \sum_{i=1}^{N} \begin{pmatrix} Z_i\{Y_i + w(Y_i - 1)\}\dot{g}\{(2Y_i - 1)(\beta' Z_i - \tilde{d})\} \\ \{Y_i + w(Y_i - 1)\}\dot{g}\{(2Y_i - 1)(\beta' Z_i - \tilde{d})\} \\ \{1 + \exp(-\tilde{d} + \beta' Z_i)\}^{-1}(1 - Y_i)/\rho - \pi_0 \\ I(\beta' Z_i - d < 0)(1 - Y_i)/\rho - \pi_0 \end{pmatrix}$$

$$= (1 - \rho) N_1^{-1} \sum_{Y_i = 1} \begin{pmatrix} Z_i \dot{g}(\beta' Z_i - \tilde{d}) \\ \dot{g}(\beta' Z_i - \tilde{d}) \\ 0 \\ 0 \end{pmatrix} + \rho N_0^{-1} \sum_{Y_i = 0} \begin{pmatrix} -w Z_i \dot{g}(\tilde{d} - \beta' Z_i) \\ -w \dot{g}(\tilde{d} - \beta' Z_i) \\ \{1 + \exp(-\tilde{d} + \beta' Z_i)\}^{-1} - \pi_0 \\ I(\beta' Z_i - d < 0) - \pi_0 \end{pmatrix}.$$

Now consider the classes of functions

$$\mathcal{G}_1 = \{I(\beta' z \le d) \mid \beta \in \Omega_\beta, d \in \Omega_d\}, \quad \mathcal{G}_2 = \left\{ \frac{1}{1 + e^{\beta' z - d}} \;\middle|\; \beta \in \Omega_\beta, d \in \Omega_d \right\},$$

$$\mathcal{G}_3 = \left\{ \frac{z}{1 + e^{\beta' z - d}} \;\middle|\; \beta \in \Omega_\beta, d \in \Omega_d \right\}, \quad \mathcal{G}_4 = \left\{ \frac{w}{1 + e^{d - \beta' z}} \;\middle|\; \beta \in \Omega_\beta, d \in \Omega_d, w \in \Omega_w \right\},$$

and

$$\mathcal{G}_5 = \left\{ \frac{wz}{1 + e^{d - \beta' z}} \;\middle|\; \beta \in \Omega_\beta, d \in \Omega_d, w \in \Omega_w \right\}.$$

The class of functions of $\{\beta' z - d | \beta, d\}$ is VC-class and thus $\mathcal{G}_1$ is both $P-$Glivenko-Cantelli and Donsker. Since $\phi(x) = \{1 + \exp(x)\}^{-1}$ is continuous and uniformly bounded by 1, $\mathcal{G}_2$ is $P-$Glivenko-Cantelli as well by the preservation property of $P-$Glivenko-Cantelli [Kosorok, 2006]. Similarly, one may verify that $\mathcal{G}_j, j = 3, 4, 5$ are $P-$Glivenko-Cantelli. Since $\phi(x)$ is

Lipschitz continuous, $\mathcal{G}_2$ is Donsker as well. Therefore, by the uniform law of large numbers,

$$\sup_{(\beta',\tilde{d},d,w)'\in\Omega} |S_N(\beta,\tilde{d},d,w) - s_0(\beta,\tilde{d},d,w)| = o_p(1),$$

which implies the consistency of $(\hat{\beta}'_S, \hat{\tilde{d}}_S, \hat{d}_S, \hat{w}_S)'$, the solution to the estimating equation $S_N(\beta,\tilde{d},d,w) = o_p(N^{-1/2})$ [Kosorok, 2006].

Furthermore, one may show that all the classes of functions $\mathcal{G}_j, j = 1,\cdots,5$ are Donsker as well, which coupled with the consistency of $(\hat{\beta}'_S, \hat{\tilde{d}}_S, \hat{d}_S, \hat{w}_S)'$ implies that

$$\sqrt{N}\{S_N(\hat{\beta}_S, \hat{\tilde{d}}_S, \hat{d}_S, \hat{w}_S) - s_0(\hat{\beta}_S, \hat{\tilde{d}}_S, \hat{d}_S, \hat{w}_S) - S_N(\beta_S, \tilde{d}_S, d_S, w_S) + s_0(\beta_S, \tilde{d}_S, d_S, w_S)\} = o_p(1).$$

Therefore

$$A_0\sqrt{N}\begin{pmatrix} \hat{\beta}_S - \beta_S \\ \hat{\tilde{d}}_S - \tilde{d}_S \\ \hat{d}_S - d_S \\ \hat{w}_S - w_S \end{pmatrix} = -\sqrt{N}S_N(\beta_S, \tilde{d}_S, d_S, w_S) + o_P(1).$$

By central limit theorem, $\sqrt{N}S_N(\beta_S, \tilde{d}_S, d_S, w_S)$ converges weakly to a mean zero Gaussian distribution with a variance-covariance matrix of

$$B_0 = (1-\rho)E\left[\begin{pmatrix} Z_i\dot{g}(\beta'_S Z_i - \tilde{d}_S) \\ \dot{g}(\beta'_S Z_i - \tilde{d}_S) \\ 0 \\ 0 \end{pmatrix}^{\otimes 2} \Bigg| Y_i = 1\right] + \rho E\left[\begin{pmatrix} -wZ_i\dot{g}(\tilde{d}_S - \beta'_S Z_i) \\ -w\dot{g}(\tilde{d}_S - \beta'_S Z_i) \\ \{1+\exp(-\tilde{d}_S + \beta'_S Z_i)\}^{-1} - \pi_0 \\ I(\beta'_S Z_i - d_S < 0) - \pi_0 \end{pmatrix}^{\otimes 2} \Bigg| Y_i = 0\right].$$

Therefore, by Sluscky theorem, $\sqrt{N}(\hat{\beta}_S' - \beta_S', \hat{\tilde{d}}_S - \tilde{d}_S, \hat{d}_S - d_S, \hat{w}_S - w_S)$ converges weakly to $N(0, A_0^{-1}B_0(A_0^{-1})')$.

To justify the resampling method, noting the fact that $(\beta_S^{*\prime}, \tilde{d}_S^*, d_S^*, w_S^*)'$ is the root of the estimating equation

$$S_N^*(\beta, \tilde{d}, d, w) = N^{-1}\sum_{i=1}^{N} B_i \begin{pmatrix} Z_i\{Y_i + w(Y_i - 1)\}\dot{g}\{(2Y_i - 1)(\beta' Z_i - \tilde{d})\} \\ \{Y_i + w(Y_i - 1)\}\dot{g}\{(2Y_i - 1)(\beta' Z_i - \tilde{d})\} \\ \{1 + \exp(-\tilde{d} + \beta' Z_i)\}^{-1}(1 - Y_i)/\rho - \pi_0 \\ I(\beta' Z_i - d < 0)(1 - Y_i)/\rho - \pi_0 \end{pmatrix} = 0,$$

we may use the similar arguments to show that

$$A_0\sqrt{N}\begin{pmatrix} \beta_S^* - \hat{\beta}_S \\ \tilde{d}_S^* - \hat{\tilde{d}}_S \\ d_S^* - \hat{d}_S \\ w_S^* - \hat{w}_S \end{pmatrix} = -\sqrt{N}S_N^*(\hat{\beta}_S, \hat{\tilde{d}}_S, \hat{d}_S, \hat{w}_S) + o_{p^*}(1).$$

By central limit theorem, conditional on the observed data, $\sqrt{N}S_N^*(\hat{\beta}_S, \hat{\tilde{d}}_S, \hat{d}_S, \hat{w}_S)$ converges weakly to a mean zero Gaussian distribution with a variance-covariance matrix B_0 as $N \to \infty$. Therefore

$$\sup_x \left| P\left\{ \sqrt{N}\begin{pmatrix} \beta_S^* - \hat{\beta}_S \\ \tilde{d}_S^* - \hat{\tilde{d}}_S \\ d_S^* - \hat{d}_S \\ w_S^* - \hat{w}_S \end{pmatrix} \le x \,\middle|\, (Y_i, X_i), i = 1, \cdots, N \right) - P\{N(0, A_0^{-1}B_0(A_0^{-1})') \le x\} \right| = o_p(1),$$

and one may use the conditional variance of $\sqrt{N}(\beta_S^* - \hat{\beta}_S)$ to approximate that of $\sqrt{N}(\hat{\beta}_S - \beta_S)$ when N is large.

For inferences on the sensitivity and specificity in a future population, we note that the classes of functions $\{I(\beta' z - d \geq 0) | \beta \in \Omega_\beta, d \in \Omega_d\}$ and $\{I(\beta' z - d \leq 0) | \beta \in \Omega_\beta, d \in \Omega_d\}$ are VC-class and thus Donsker. Coupled with the root-N consistency of $\hat{\beta}_S$ and $\hat{d}_S$, the stochastic continuity associated with the Donsker property suggests that

$$\begin{bmatrix} N_1^{-1/2} \sum_{Y_i=1} \left\{ I(\hat{\beta}_S' Z_i \geq \hat{d}_S) - P(\hat{\beta}_S' Z_i \geq \hat{d}_S \mid Y_i = 1, \hat{\beta}_S, \hat{d}_S) \right\} \\ N_0^{-1/2} \sum_{Y_i=0} \left\{ I(\hat{\beta}_S' Z_i \leq \hat{d}_S) - P(\hat{\beta}_S' Z_i \leq \hat{d}_S \mid Y_i = 0, \hat{\beta}_S, \hat{d}_S) \right\} \end{bmatrix}$$

$$= \begin{bmatrix} N_1^{-1/2} \sum_{Y_i=1} \{ I(\beta_S' Z_i \geq d_S) - P(\beta_S' Z_i \geq d_S \mid Y_i = 1) \} \\ N_0^{-1/2} \sum_{Y_i=0} \{ I(\beta_S' Z_i \leq d_S) - P(\beta_S' Z_i \leq d_S \mid Y_i = 0) \} \end{bmatrix} + o_p(1).$$

Therefore, the asymptotical validity of $I_{\alpha,\eta}$ and $I_{\alpha,\pi}$ follows from the fact that $\hat{\eta}_0 - \eta_0 = o_p(1)$ and

$$\frac{1}{N_0} \sum_{Y_i=0} I(\hat{\beta}_S' Z_i \leq \hat{d}_S) = \pi_0.$$

Appendix C. Asymptotical Properties for $\hat{\beta}_R$ and $\hat{d}_R(\cdot)$

To study the asymptotical property of $\hat{\beta}_R$ and $\hat{d}_R(\cdot)$, we need the following notations and regularity conditions:

1. The unknown parameters of the model are β, $\tilde{d}(\pi)$, $d(\pi)$ and $w(\pi), \pi \in [\pi_L, \pi_U]$. Without of loss of generality, we assume that the parameter space is $\Omega = \Omega_\beta \times B[\pi_L, \pi_U]^{\otimes 3}$, where $B[\pi_L, \pi_U]$ is the class of all functions on $[\pi_L, \pi_U]$ uniformly bounded by a large constant C_0.

2. For any given positive weight function $w(\cdot) \in B[\pi_L, \pi_U]$, the functional estimating equation

$$\tilde{m}_0\{\beta, d(\cdot), w(\cdot)\}(\pi) = 0, \pi \in [\pi_L, \pi_U]$$

has an unique solution β_w and $d_w(\cdot) \in B[\pi_L, \pi_U]$, where $\tilde{m}_0 : \Omega \to R^p \times l^\infty[\pi_L, \pi_U]$ is the functional: $\tilde{m}_0\{\beta, d(\cdot), w(\cdot)\}(\pi) =$

$$(1-\rho)E\left(\begin{array}{c} \int_{\pi_L}^{\pi_U} Z_i \dot{g}\{\beta' Z_i - d(\pi)\}d\pi \\ \dot{g}\{\beta' Z_i - d(\pi)\} \end{array}\middle| Y_i = 1\right) - \rho E\left(\begin{array}{c} \int_{\pi_L}^{\pi_U} Z_i w(\pi) \dot{g}\{-\beta' Z_i + d(\pi)\}d\pi \\ w(\pi)\dot{g}\{-\beta' Z_i + d(\pi)\} \end{array}\middle| Y_i = 0\right).$$

3. $E\{[1+\exp\{\beta_w' Z_i - d_w(\pi)\}]^{-1} | Y_i = 0\} = \pi, \pi \in [\pi_L, \pi_U]$ has an unique solution for $w(\cdot) \in B_{C_0}[\pi_L, \pi_U]$.

These conditions ensure the existence of the unique solution to the functional estimating equation

$$m_0\{\beta, \tilde{d}(\cdot), d(\cdot), w(\cdot)\}(\pi) = \begin{pmatrix} \tilde{m}_0\{\beta, \tilde{d}(\cdot), w(\cdot)\}(\pi) \\ E\{[1+\exp\{\beta' Z_i - \tilde{d}(\pi)\}]^{-1} | Y_i = 0\} - \pi \\ P\{\beta' Z_i \le d(\pi) | Y_i = 0\} - \pi \end{pmatrix} = 0, \qquad \text{(S0.2)}$$

where $\pi \in [\pi_L, \pi_U]$. Let $\beta_R, \tilde{d}_R(\cdot), d_R(\cdot)$ and $w_R(\cdot)$ denote the root. We further assume that

4 The linear functional $\dot{m}_{\beta_R,\tilde{d}_R,d_R,w_R}\{\beta,\tilde{d}(\cdot),d(\cdot),w(\cdot)\}(\pi):\Omega\to R^p\times l^\infty[\pi_L,\pi_U]^{\otimes 3}$

$$(1-\rho)\begin{pmatrix} \int_{\pi_L}^{\pi_U} E[\{Z_i^{\otimes 2}\beta-\tilde{d}(\pi)Z_i\}\ddot{g}\{\beta_R'Z_i-\tilde{d}_R(\pi)\}]d\pi \\ E[\{\beta'Z_i-\tilde{d}(\pi)\}\ddot{g}\{\beta_R'Z_i-\tilde{d}_R(\pi)\}] \\ 0 \\ 0 \end{pmatrix}\Bigg| Y_i=1$$

$$+\rho\begin{pmatrix} \int_{\pi_L}^{\pi_U} E\left[\{Z_i^{\otimes 2}\beta-\tilde{d}(\pi)Z_i\}\ddot{g}\{-\beta_R'Z_i+\tilde{d}_R(\pi)\}w_R(\pi)-Z_i\dot{g}\{-\beta_R'Z_i+\tilde{d}_R(\pi)\}w(\pi)\right]d\pi \\ E\left[\{\beta'Z_i-\tilde{d}(\pi)\}\ddot{g}\{-\beta_R'Z_i+\tilde{d}_R(\pi)\}w_R(\pi)-Z_i\dot{g}\{-\beta_R'Z_i+\tilde{d}_R(\pi)\}w(\pi)\right] \\ -E\left(\{\beta'Z_i+\tilde{d}(\pi)\}\exp\{\beta_R'Z_i-\tilde{d}_R(\pi)\}[1+\exp\{\beta_R'Z_i-\tilde{d}_R(\pi)\}]^{-2}\right) \\ \beta'\dot{P}_\beta\{\beta_R,d_R(\pi)\}+d(\pi)\dot{P}_d\{\beta_R,d_R(\pi)\} \end{pmatrix}\Bigg| Y_i=0$$

is continuously invertible, where $\dot{P}_\beta\{\beta_R,d_R(\pi)\}$ and $\dot{P}_d\{\beta_R,d_R(\pi)\}$ are partial derivatives of $P(\beta'Z_i\le d|Y_i=0)$ with respect to β and d, respectively.

5 $M_N\{\beta,\tilde{d}(\cdot),d(\cdot),w(\cdot)\}:\Omega\to R^p\times l^\infty[\pi_L,\pi_U]^{\otimes 3}$ is the functional: $M_N\{\beta,\tilde{d}(\cdot),d(\cdot),w(\cdot)\}(\pi)=$

$$(1-\rho)N_1^{-1}\sum_{Y_i=1}\begin{pmatrix} \int_{\pi_L}^{\pi_U} Z_i\dot{g}\{\beta'Z_i-d(\pi)\}d\pi \\ \dot{g}\{\beta'Z_i-d(\pi)\} \\ 0 \\ 0 \end{pmatrix}+\rho N_0^{-1}\sum_{Y_i=0}\begin{pmatrix} -\int_{\pi_L}^{\pi_U} Z_i w(\pi)\dot{g}\{-\beta'Z_i+d(\pi)\}d\pi \\ -w(\pi)\dot{g}\{-\beta'Z_i+d(\pi)\} \\ \left[1+\exp\{\beta'Z_i-\tilde{d}(\pi)\}\right]^{-1}-\pi \\ I\{\beta'Z_i\le d(\pi)\}-\pi \end{pmatrix},$$

and $\{\hat{\beta}_R',\hat{\tilde{d}}_R(\cdot),\hat{d}_R(\cdot),\hat{w}_R(\cdot)\}'$ is the root of the estimating function

$$M_N\{\beta,\tilde{d}(\cdot),d(\cdot),w(\cdot)\}(\pi)=0,\pi\in[\pi_L,\pi_U].$$

Firstly since

$$\Big| S_N(\beta,\tilde d,d,w) - s_0(\beta,\tilde d,d,w) \Big| = o_p(1)$$

uniformly over any compact set of $(\beta,\tilde d,w)$,

$$\begin{aligned}
&\sup_{\Omega}\Bigg| N_1^{-1}\sum_{Y_i=1}\int_{\pi_L}^{\pi_U} Z_i\dot g\{\beta' Z_i-\tilde d(\pi)\}d\pi - \int_{\pi_L}^{\pi_U} E\left(Z_i\dot g\{\beta' Z_i-\tilde d(\pi)\}\mid Y_i=1\right)d\pi\Bigg|\\
\le&\sup_{\Omega}\int_{\pi_L}^{\pi_U}\Bigg| N_1^{-1}\sum_{Y_i=1} Z_i\dot g\{\beta' Z_i-\tilde d(\pi)\} - E\left(Z_i\dot g\{\beta' Z_i-\tilde d(\pi)\}\mid Y_i=1\right)\Bigg| d\pi\\
\le&(\pi_U-\pi_L)\sup_{\beta,\tilde d}\Bigg| N_1^{-1}\sum_{Y_i=1} Z_i\dot g(\beta' Z_i-\tilde d) - E\left(Z_i\dot g(\beta' Z_i-\tilde d)\mid Y_i=1\right)\Bigg|\\
=&o_p(1).
\end{aligned}$$

Similarly

$$\begin{aligned}
&\sup_{\Omega}\Bigg| N_0^{-1}\sum_{Y_i=0}\int_{\pi_L}^{\pi_U} Z_i w(\pi)\dot g\{-\beta' Z_i+\tilde d(\pi)\}d\pi - \int_{\pi_L}^{\pi_U} E\left(Z_i w(\pi)\dot g\{-\beta' Z_i+\tilde d(\pi)\}\mid Y_i=0\right)d\pi\Bigg|\\
\le&\sup_{\Omega}\int_{\pi_L}^{\pi_U}\Bigg| N_0^{-1}\sum_{Y_i=0} Z_i w(\pi)\dot g\{-\beta' Z_i+\tilde d(\pi)\} - E\left(Z_i w(\pi)\dot g\{-\beta' Z_i+\tilde d(\pi)\}\mid Y_i=0\right)\Bigg| d\pi\\
\le&(\pi_U-\pi_L)\sup_{\beta,\tilde d}\Bigg| N_0^{-1}\sum_{Y_i=0} Z_i\dot g(-\beta' Z_i+\tilde d) - E\left(Z_i\dot g(-\beta' Z_i+\tilde d)\mid Y_i=0\right)\Bigg| \sup_w |w|\\
=&o_p(1).
\end{aligned}$$

One can also can show that

$$\sup_{\Omega\times[\pi_L,\pi_U]}\Bigg| N_1^{-1}\sum_{Y_i=1}\dot g\{\beta' Z_i-\tilde d(\pi)\} - E\left(\dot g\{\beta' Z_i-\tilde d(\pi)\}\mid Y_i=1\right)\Bigg| = o_p(1),$$

$$\sup_{\Omega\times[\pi_L,\pi_U]} \left| N_0^{-1}\sum_{Y_i=0}^{N} w(\pi)\dot{g}\{-\beta' Z_i+\tilde{d}(\pi)\}]-E\left(w(\pi)\dot{g}\{-\beta' Z_i+\tilde{d}(\pi)\}\right)\right| = o_p(1),$$

$$\sup_{\Omega\times[\pi_L,\pi_U]} \left| N_0^{-1}\sum_{Y_i=0}^{N}(1-Y_i)\left[1+\exp\{\beta' Z_i-\tilde{d}(\pi)\}\right]^{-1}-E\left(\left[1+\exp\{\beta' Z_i-\tilde{d}(\pi)\}\right]^{-1}\big|Y_i=0\right)\right| = o_p(1),$$

and

$$\sup_{\Omega\times[\pi_L,\pi_U]} \left| N_0^{-1}\sum_{Y_i=0} I\{\beta' Z_i-d(\pi)<0\}-P\{\beta' Z_i<d(\pi)\mid Y_i=0\}\right| = o_p(1).$$

Thus we have

$$\sup_{(\beta',\tilde{d}(\cdot),d(\cdot),w(\cdot))'\in\Omega,\pi\in[\pi_L,\pi_U]} |M_N\{\beta,\tilde{d}(\cdot),d(\cdot),w(\cdot)\}(\pi)-m_0\{\beta,\tilde{d}(\cdot),d(\cdot),w(\cdot)\}(\pi)| = o_p(1).$$

Coupled with the uniqueness of the solution for the estimating equation $m_0\{\beta,\tilde{d}(\cdot),d(\cdot),w(\cdot)\}(\cdot)=0$, this uniform convergence implies that

$$|\hat{\beta}_R-\beta_R|+\sup_{\pi\in[\pi_L,\pi_U]}\{|\hat{\tilde{d}}_R(\pi)-\tilde{d}_R(\pi)|+|\hat{d}_R(\pi)-d_R(\pi)|+|\hat{w}_R(\pi)-w_R(\pi)|\} = o_p(1).$$

One can verify that $m\{\beta,\tilde{d}(\cdot),d(\cdot),w(\cdot)\}$ is uniformly Frechet-differentiable with the continuously invertible derivative of $\dot{m}_{\beta_R,\tilde{d}_R,d_R,w_R}\{\beta,\tilde{d}(\cdot),d(\cdot),w(\cdot)\}$ at $(\beta_R',\tilde{d}_R,d_R,w_R)'$. Furthermore, since the process

$$\sqrt{N}\{S_N(\theta)-s_0(\theta)\}$$

indexed by $\theta=(\beta',\tilde{d},d,w)'\in\Omega_0$ is asymptotical tight and the natural metric induced by the process is equivalent to $|\theta_2 - \theta_1|$, we have

$$\lim_{\delta\to 0}\limsup_{N\to\infty} P\left(\sup_{|\theta_2-\theta_1|\le\delta}\left|\sqrt{N}\{S_N(\theta_2)-s_0(\theta_2)\}-\sqrt{N}\{S_N(\theta_1)-s_0(\theta_1)\}\right|\ge\delta\right)=0.$$

Furthermore, since

$$\sup_{\substack{\pi\in[\pi_L,\pi_U],\,|\beta_2-\beta_1|+\\ \sup_\pi|\tilde d_2(\pi)-\tilde d_1(\pi)|+\\ \sup_\pi|d_2(\pi)-d_1(\pi)|+\\ \sup_\pi|w_2(\pi)-w_1(\pi)|\le\delta}}\left|\sqrt{N}(M_N-m_0)\{\beta_2,\tilde d_2(\cdot),d_2(\cdot),w_2(\cdot)\}(\pi)-\sqrt{N}(M_N-m_0)\{\beta_1,\tilde d_1(\cdot),d_1(\cdot),w_1(\cdot)\}(\pi)\right|$$

$$\le C\sup_{|\theta_2-\theta_1|\le\delta}\left|\sqrt{N}\{S_N(\theta_2)-s_0(\theta_2)\}-\sqrt{N}\{S_N(\theta_1)-s_0(\theta_1)\}\right|$$

for a positive constant C, $\sqrt{N}(M_N-m_0)\{\beta,\tilde d(\cdot),d(\cdot),w(\cdot)\}$ is also asymptotically tight and we have

$$\sup_{\pi\in[\pi_L,\pi_U]}\left|\sqrt{N}(M_N-m_0)\{\hat\beta_R,\hat{\tilde d}_R(\cdot),\hat d_R(\cdot),\hat w_R(\cdot)\}(\pi)-\sqrt{N}(M_N-m_0)\{\beta_R,\tilde d_R(\cdot),d_R(\cdot),w_R(\cdot)\}(\pi)\right|=o_p(1),$$

due to the consistency of the estimator $(\hat\beta_R', \hat{\tilde d}_R, \hat d_R(\cdot), \hat w_R(\cdot))'$. Therefore, we have

$$\sqrt{N}\begin{pmatrix}\hat\beta_R-\beta_R\\ \hat{\tilde d}_R(\cdot)-\tilde d_R(\cdot)\\ \hat d_R(\cdot)-d_R(\cdot)\\ \hat w_R(\cdot)-w_R(\cdot)\end{pmatrix}=-\dot m_0^{-1}\left(\sqrt{N}M_N\{\beta_R,\tilde d_R(\cdot),d_R(\cdot),w_R(\cdot)\}\right)+o_p(1)$$

Lastly, since we assume that $\tilde d_R(\cdot)$, $d_R(\cdot)$ and $w_R(\cdot)$ are uniformly continuous on $[\pi_L,\pi_U]$, the process $\sqrt{N}M_N\{\beta_R,\tilde d_R(\cdot),d_R(\cdot),w_R(\cdot)\}(\pi)$ is asymptotically tight and thus by the functional central limit theorem,

$$\sqrt{N}M_N\{\beta_R,\tilde d_R(\cdot),d_R(\cdot),w_R(\cdot)\}(\pi),\pi\in[\pi_L,\pi_U]$$

weakly converges to a mean zero Gaussian process and especially $\sqrt{N}(\hat{\beta}_R - \beta_R)$ converges to a multivariate mean zero Gaussian in distribution.

Appendix D. ROC curves from the optimal discriminant function and standard logistic regression.

To summarize various cases used in the simulation study, we plotted the ROC curves based on the optimal discriminant function and the linear combination from fitting the simple logistic regression. The purpose is to demonstrate the potential space improvement space for the standard logistic regression in each of the eight cases. To this end, the optimal discriminant function is defined as the ratio of the density function of cases and that of controls. The weight used in the linear combination from the standard logistic regression is obtained by averaging the estimated regression coefficients from 10,000 simulated data sets. It is clear that while the standard logistic regression is equivalent to the optimal discriminant function in case 1, it is far from optimal for cases 2-8, and, in particular, for cases 2 and 8. It is conceivable that linear combination different from that of the standard logistic regression may better approximate the nonlinear optimal decision boundary at selected specificity levels. These simulation designs are thus interesting cases to study whether the proposed method can improve the standard logistic regression.

Appendix E. Simulation study for the area under the partial ROC curve.

We investigated the parallel properties of partial ROC curve-based combinations with the same simulation settings described in the main paper.

In the first set of simulations, we examined the realized area under the partial ROC curve in settings 1-8 described in the main paper. In this simulation, we let $[\pi_L, \pi_U] = [0.85, 0.95]$, over which the area under the partial ROC curve is of our interest. The results are summarized in Figure 3. In case 1, where the logistic regression is the true model, the proposed method and logistic regression perform similarly and are better than the grid search method. In cases 2-4, while the grid search directly maximizing the area under the partial ROC curve yields the biggest partial area, the performance of the proposed method is comparable and clearly superior to that of the logistic regression. In cases 5-8, the proposed new method also yields substantially bigger areas under the partial ROC curve than the logistic regression.

In the second and third sets of simulations, we examined the empirical performance of the resampling method for $\hat{\beta}_R$ and true coverage level of the credible sets of the area under the partial ROC curve in the same settings as those studied in the second and third sets of simulations in the main paper, respectively. The results are reported in Table 1 and Figure 4. In general, both the confidence interval for β_R and the credible set for the true area under the partial ROC curve have achieved satisfactory empirical coverage level, supporting the validity of the proposed inference procedure.

Lastly, in the forth set of simulations, we investigated the ability of the ROC-based ensemble method for correctly identifying the informative features. The simulation design is the same as that in the fourth set of the simulations of the main paper. We have applied the logistic regression as well as proposed method aiming that maximizes the AUC under the partial ROC curve corresponding to specificity levels between 85% and 95%. The results are reported in Table 2. Both methods can identify important features with high probability and the logistic regression tends to select more informative as well as noise features than the new method. The true areas under the partial ROC curves from these two methods are similar.

Table 1: Simulation results for evaluating the empirical performance of the resampling method based on 500 simulations: bias, empirical bias; ESE, empirical standard error; ASE, empirical average of the estimated standard error; COV, the empirical coverage probability.

	ROC-based ensemble				
case	β_R	bias	ESE	ASE	COV
3	-0.109	-0.018	0.137	0.125	92.0%
	0.148	0.012	0.137	0.127	92.8%
	1.051	0.028	0.137	0.149	93.6%
4	0.933	0.021	0.163	0.149	94.0%
	0.387	0.027	0.183	0.162	93.0%
	0.929	0.024	0.159	0.149	94.2%
6	0.360	0.029	0.221	0.218	93.8%
	1.496	0.036	0.274	0.256	92.8%
	0.703	0.032	0.195	0.193	94.8%
	-1.232	-0.008	0.196	0.195	94.6%
7	0.360	0.029	0.221	0.218	93.8%
	1.496	0.036	0.274	0.256	92.8%
	0.703	0.032	0.195	0.193	94.8%
	-1.232	-0.008	0.196	0.195	94.6%

Table 2: The empirical probabilities of selecting informative as well as noise features based on lasso-regularized logistic regression and the proposed ROC-based ensemble method. The average area under the partial ROC curve (AUC) is also reported.

method		Emp. Prob. of Being Selected					AUC
	ρ	Z_1	Z_2	Z_3	Z_4	Noise markers	
Logistic reg.		100%	100%	99%	72%	13%	0.056
ROC-based		100%	100%	90%	41%	5%	0.056
Logistic reg.		100%	100%	100%	74%	11%	0.073
ROC-based		100%	100%	93%	45%	3%	0.073
Logistic reg.		100%	100%	91%	65%	9%	0.086
ROC-based		96%	100%	80%	36%	3%	0.087

Bibliography

M. Kosorok. *Introduction ot empirical processes and semiparametric inference.* Springer-New York, 2006.

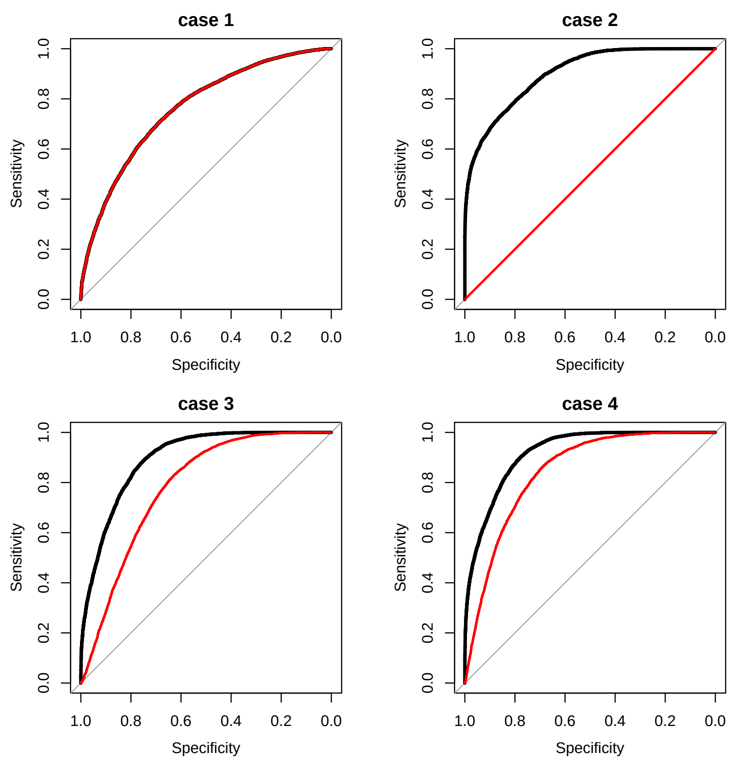

Figure 1: ROC curves based on the optimal discriminant function and the linear combination from fitting the standard logistic regression for cases 1-4: red, standard logistic regression; black, optimal discriminant function.

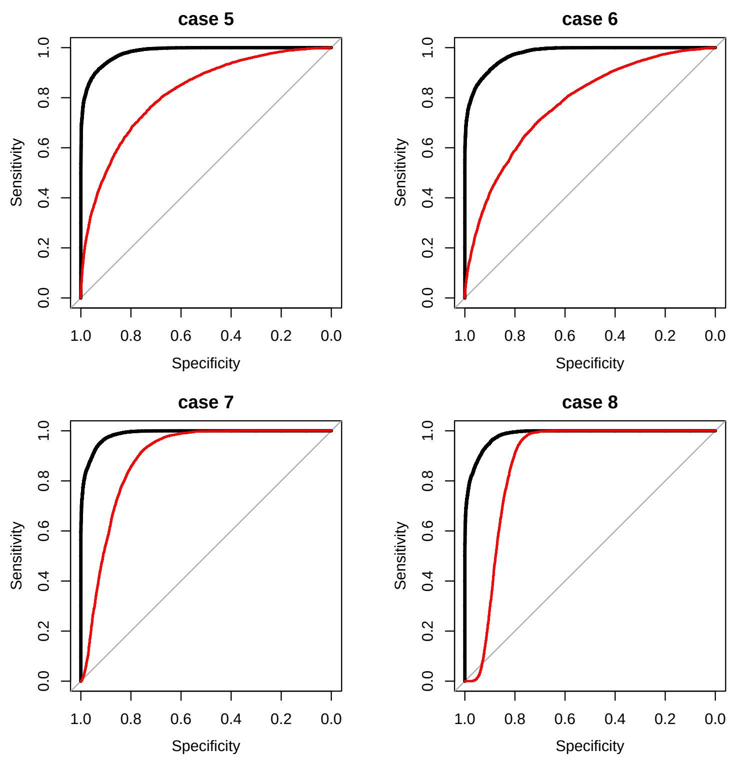

Figure 2: ROC curves based on the optimal discriminant function and the linear combination from fitting the standard logistic regression for cases 5-8: red, standard logistic regression; black, optimal discriminant function.

Figure 3: Boxplots for the empirical distributions of the realized area under the partial ROC curve of the risk score constructed using three different methods: white, logistic regression; light gray, new proposal; dark gray, grid search.

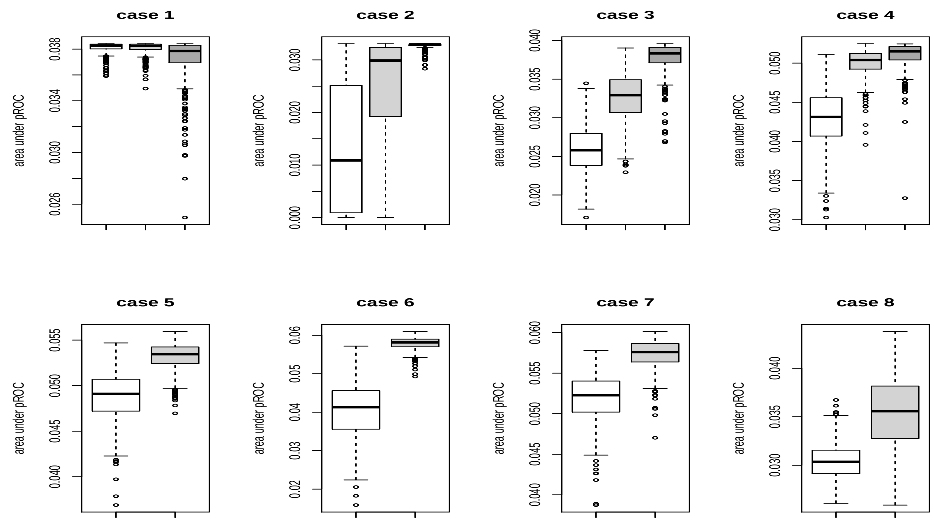

Figure 4: Empirical coverage levels for the constructed credible sets of the area under the partial ROC curve