Generalization of t statistic and AUC by considering heterogeneity in probability distributions

Osamu Komori Department of Mathematical Analysis and Statistical Inference, Project Assistant Professor

1 Generalized AUC

We discuss a statistical method of a classification problem for two groups. For a binary class label $y \in \{0,1\}$ and a covariate vector $x \in \mathbb{R}^p$, we consider a statistical situation in which the neither conditional distribution of x given $y=0$ nor given $y=1$ are well modelled by a specific distribution.

For a sample $\{x_{0i}:\ i=1,\dots,n_0\}$ for $y=0$ and a sample $\{x_{1j}:\ j=1,\dots,n_1\}$ for $y=1$ where $n=n_0+n_1$, we propose a generalized u-statistic defined by

$$L_U(\beta)=\frac{1}{n_0 n_1}\sum_{i=1}^{n_0}\sum_{j=1}^{n_1}U\left\{\frac{\beta^{\mathrm{T}}(x_{1j}-x_{0i})}{(\beta^{\mathrm{T}}S\beta)^{1/2}}\right\}, \tag{1}$$

where U is an arbitrary real-valued function: $\mathbb{R}\to\mathbb{R}$; S is a normalizing factor given as

$$S=\frac{1}{n}\sum_{i=1}^{n_0}(x_{0i}-\bar{x}_0)(x_{0i}-\bar{x}_0)^\top+\frac{1}{n}\sum_{j=1}^{n_1}(x_{1j}-\bar{x}_1)(x_{1j}-\bar{x}_1)^\top. \tag{2}$$

2 Asymptotic consistency and normality

Let us consider the estimator associated with the generalized t-statistic as

$$\widehat{\beta}_U=\underset{\beta\in\mathbb{R}^p}{\mathrm{argmax}}\ L_U(\beta). \tag{3}$$

Then we consider the following assumption:

(A) $\qquad E_y(g_y \mid w_y=a)=0 \quad$ for all $a\in\mathbb{R}$, for $y=0,1$

where $w_y=\beta_0^{\mathrm{T}}x_y$, $g_y=Qx_y$, $Q=I-\Sigma\beta_0\beta_0^{\mathrm{T}}$, $\Sigma_y^*=Q\Sigma_yQ^{\mathrm{T}}$, $\mu_0+\mu_1=0$, and

$$\beta_0=\frac{\Sigma^{-1}(\mu_1-\mu_0)}{\{(\mu_1-\mu_0)^{\mathrm{T}}\Sigma^{-1}(\mu_1-\mu_0)\}^{1/2}}. \tag{4}$$

Theorem 2.1 *Under Assumption (A), $\widehat{\beta}_U$ is asymptotically consistent with β_0 for any U.*

Next we consider the following assumption in addition to (A):

(B) $\qquad \mathrm{var}_y(g_y \mid w_y=a)=\Sigma_y^* \quad$ for all $a\in\mathbb{R}$, for $y=0,1$

where var_y denotes the conditional variance of x given y. Then we assume mixture model for class label $y\in\{0,1\}$.

$$p_y(x)\ =\sum_{k=1}^{\infty}\epsilon_{yk}\phi(x,\nu_{yk},V_{yk}). \tag{5}$$

Theorem 2.2 *For $y=0,1$ assumptions (A) and (B) under the infinite mixture model in (5) are equivalent to*

(A′) $\displaystyle\sum_{k\in K_{y\ell}}\epsilon_k(Q-Q_{yk})=0,\quad \sum_{k\in K_{y\ell}}\epsilon_{yk}Q_{yk}\nu_{yk}=0,\ \text{for } {}^\forall\ell\in\mathbb{N},\ y=0,1$

(B′) $\displaystyle\sum_{k\in K_{y\ell}}\epsilon_{yk}\Big\{Q_{yk}V_{yk}Q-Q\Sigma_yQ\Big\}=0,\ \text{for } {}^\forall\ell\in\mathbb{N}, y=0,1$

where $Q_{yk}=I_p-V_{yk}\beta^*\beta^{*\top}/(\beta^{*\top}V_{yk}\beta^*)$, $K_{y\ell}=\{k\ \mid\ \beta^{*\top}\nu_{yk}=\beta^{*\top}\nu_{y\ell},\ \beta^{*\top}V_{yk}\beta^*=\beta^{*\top}V_{y\ell}\beta^*\}$.

Here we assume the following semiparametric model for probability density functions,

$$p_y(x)=\psi_y(c+\beta^\top x)(2\pi)^{-\frac{p}{2}}|\Sigma_y|^{-\frac{1}{2}}\exp\Big(-\frac{x^\top\Sigma_y^{-1}x}{2}\Big),\ \text{for } y=0,1, \tag{6}$$

where ψ_y is a function from $\mathbb{R}$ to $\mathbb{R}_+$ and there exists λ_y such that

$$\Sigma_y\beta=\lambda_y\beta,\ \text{for } y=0,1. \tag{7}$$

Theorem 2.3 *The target parameter β_0 is proportional to β in (6) and both assumptions (A) and (B) hold for (6).*

Theorem 2.4 *Under Assumptions (A) and (B), $n^{1/2}(\widehat{\beta}_U-\beta_0)$ is asymptotically distributed as $N(0,\Sigma_U)$, where*

$$\Sigma_U=c_U\Sigma_0^*, \tag{8}$$

$$c_U=\frac{E_0\Big[E_1\{U'(w)\}\Big]^2+E_1\Big[E_0\{U'(w)\}\Big]^2+2\rho E\{U'(w)\}E\{U'(w)w\}-\Big[E\{U'(w)w\}\Big]^2}{\Big[E\{U'(w)S(w)+U'(w)w\}\Big]^2}, \tag{9}$$

in which $S(w)=\partial\log f(w)/\partial w$, $f(w)$ is the probability density of $w=w_1-w_0$, $\rho=E(w)$ and U' denotes the first derivative of U.

3 Simulation studies

We consider normal mixtures as follows:

$$x_0\sim\epsilon_0N(\mathbf{0},\boldsymbol{I}_p)+(1-\epsilon_0)N(\boldsymbol{\nu}_0,\boldsymbol{I}_p)$$
$$x_1\sim\epsilon_1N(\boldsymbol{\nu}_1,\boldsymbol{V}_1)+\epsilon_2N(\boldsymbol{\nu}_2,\boldsymbol{V}_2)+(1-\epsilon_1-\epsilon_2)N(\boldsymbol{\nu}_3,\boldsymbol{V}_3),$$

where $\boldsymbol{\nu}_0=(-2,-0.2,\dots,-0.2)^\top$, $\boldsymbol{\nu}_1=(3,0.3,\dots,0.3)^\top$, $\boldsymbol{\nu}_2=(4,0.4,\dots,0.4)^\top$, $\boldsymbol{\nu}_3=(-1,-0.1,\dots,-0.1)^\top\in\mathbb{R}^p$, $\boldsymbol{V}_1=\boldsymbol{V}_2=\boldsymbol{V}_3=I_p$, $\epsilon_0=0.5$, $\epsilon_1=\epsilon_2=0.1$. We consider the following U functions.

1. optimal-U

$$U_{\mathrm{opt}}(w)=U_{\mathrm{upper}}(w)+a_1w+a_2w^2+\dots+a_mw^m, \tag{10}$$

where the polynomial order m is determined by the cross validation of c_U.

2. upper-U

$$U_{\mathrm{upper}}(w)=\log f(w)+\frac{1}{2}w^2-\frac{\rho^3}{2+\rho^2}w. \tag{11}$$

3. approx-U

$$U_{\mathrm{approx}}(w)=\log f(w)+\frac{\rho}{2+\rho^2}w \tag{12}$$

4. auc-U

$$U_{\mathrm{auc}}(w)=\Phi\Big(\frac{w}{\sigma}\Big), \tag{13}$$

where $\sigma=0.01$.

5. linear-U (Fisher)

$$U_{\mathrm{linear}}(w)=w \tag{14}$$

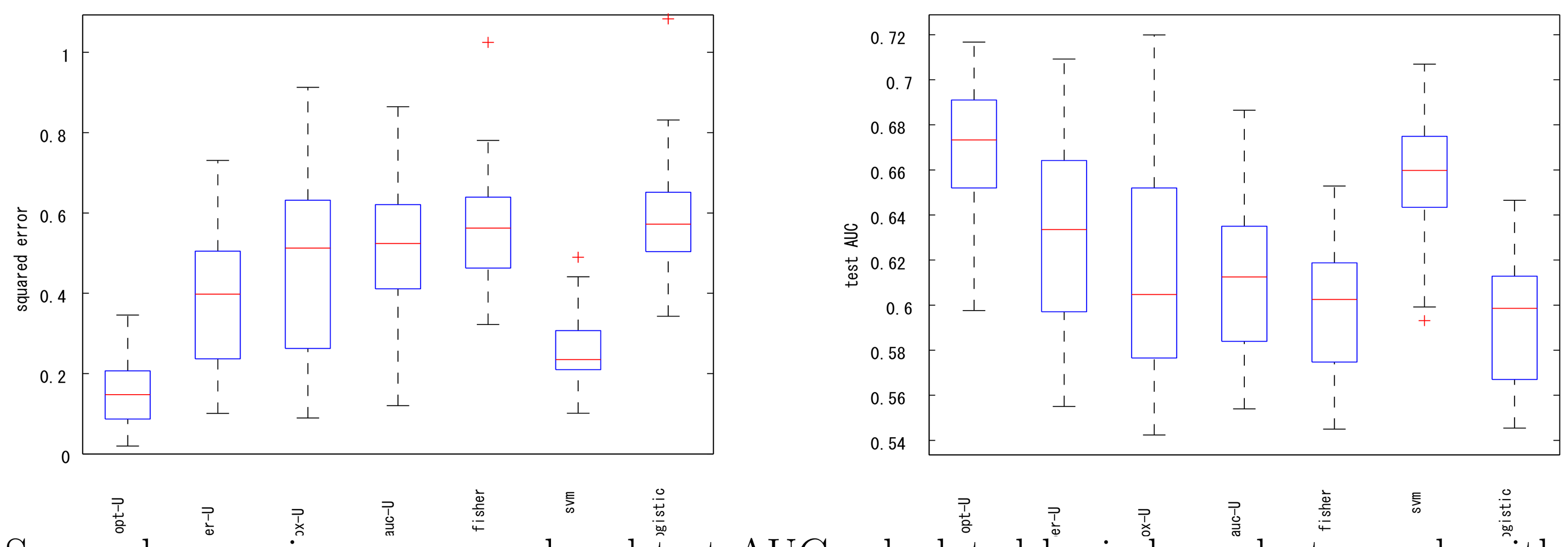

Fig1. Squared errors in upper panel and test AUC calculated by independent sample with size 1000 in lower panel, based on 30 repetitions ($p=20$ and $n_0=n_1=50$)