Outline 2

February 14, 2019

0.1 Step1:

To calculate the max purity function, first, fit the linear mixed model for the outcome $\underset{\sim}{\mathrm{y}}$ and time $\boldsymbol{X}$, with baseline covariates $\boldsymbol{x}$:

$$\boldsymbol{y} = \boldsymbol{X}(\boldsymbol{\beta} + \boldsymbol{b} + \boldsymbol{\Gamma}(\boldsymbol{\alpha}'\boldsymbol{x})) + \boldsymbol{\epsilon}.$$

We can define the covariate matrix of $\boldsymbol{X}$ as $\boldsymbol{z}$. The $\boldsymbol{z}$ contains both fixed effects and random effects.

$$\boldsymbol{z} = \beta + \boldsymbol{b} + \boldsymbol{\Gamma} x$$

We can also write the above equation in the matrix version:

$$Y = \begin{bmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ 1 & t_3 & t_3^2 \\ 1 & t_4 & t_4^2 \end{bmatrix} \left[\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix} + \begin{pmatrix} \gamma_0 \\ \gamma_1 \\ \gamma_2 \end{pmatrix} x \right] + \boldsymbol{\epsilon}$$

where
 * $\boldsymbol{\beta}$ is the vector of covariates for fixed effects of $\boldsymbol{X}$.
 * $\boldsymbol{b}$ is the vector of random effects
 * $\boldsymbol{\Gamma x}$ is the vector of fixed effects of the baseline covariates.
 * x is the combination of the input baseline covariates. $x = \boldsymbol{\alpha}'\boldsymbol{x}$ and $\boldsymbol{x}$ is the covariates vector.

0.2 Step2:

Then we need to estimate the distribution of $\underset{\sim}{z}$ for drug group and placebo, separately.

$$\begin{aligned} f(z) &= \int_x f(z, x)dx \\ &= \int_x f(z|x)g(x)dx \end{aligned}$$

where,

* the conditional distribution $f(z|x) \sim MVN(\boldsymbol{\beta} + \boldsymbol{\Gamma x}, \boldsymbol{D})$

* $g(x)$ is the distribution of the covariates combination. For example, if the covariates combination only contains "sex", which is binary, then the intergal becomes summation.

* $\boldsymbol{D}$ is the covariates matrix of random effects $\boldsymbol{b}$.

We could then fit the $f(\cdot)$ for drug group and placebo group separately, i.e. $f_1(z)$ and $f_2(z)$.

0.3 Step3:

Define the purity for a combination of covariates x as

$$f_1(z|x) = |2\pi D|^{-\frac{1}{2}} exp(-\frac{1}{2}(\boldsymbol{z} - \hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\Gamma}}_1 x)^T D^{-1}(\boldsymbol{z} - \hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\Gamma}}_1 x))$$

$$f_2(z|x) = |2\pi D|^{-\frac{1}{2}} exp(-\frac{1}{2}(\boldsymbol{z} - \hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\Gamma}}_1 x)^T D^{-1}(\boldsymbol{z} - \hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\Gamma}}_1 x))$$

$$\int_x (\frac{\pi_1 f_1(z|x) - \pi_2 f_2(z|x)}{\pi_1 f_1(z|x) + \pi_2 f_2(z|x)})^2 g(x)dx$$

which can be also approximated as the summation form:

$$\sum_x (\frac{\pi_1 f_1(z|x) - \pi_2 f_2(z|x)}{\pi_1 f_1(z|x) + \pi_2 f_2(z|x)})^2 g(x)$$

This is one purity value based on one combination of baseline covariates. Define the purity as P_x

We would like to find the $\boldsymbol{\alpha}$ that max the P_x