Outline 2

February 14, 2019

0.1 Step1:

To calculate the max purity function, first, fit the linear mixed model for the outcome $\boldsymbol{y}$ $\cancel{\underline{y}}$ 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.

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

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
- $\boldsymbol{\Gamma}$ $\cancel{\boldsymbol{x}}$ is the vector of fixed effects of the baseline covariates.
- $w = \boldsymbol{\alpha}'\boldsymbol{x}$ is the combination of the input baseline covariates.

0.2 Step 2:

~~Then we need to~~ Estimate the distribution of $\boldsymbol{z}$ for drug group and placebo, separately.

$$\begin{aligned} f(\boldsymbol{z}) &= \int_w f(\boldsymbol{z}, w) dw \\ &= \int_w f(\boldsymbol{z}|w) g(w) dw \end{aligned}$$

where,

- the conditional distribution $f(\boldsymbol{z}|w) \sim MVN(\boldsymbol{\beta} + \boldsymbol{\Gamma} w, \boldsymbol{D})$
- $g(w)$ is the distribution of the covariates combination $\boldsymbol{\alpha}'\boldsymbol{x}$. For example, if the covariates combination only contains "sex", which is binary, then the integral becomes summation.
- $\boldsymbol{D}$ is the ~~covariates~~ covariance matrix of random effects $\boldsymbol{b}$.

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

0.3 Step 3:

Define the purity for a combination of covariates $w = \boldsymbol{\alpha}'\boldsymbol{x}$ as

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

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

[Note - I fixed the exponents of the normalization constant, $p =$ dimension of $\boldsymbol{\beta}$]

$$\int_w \big[\frac{\pi_1 f_1(\boldsymbol{z}|w) - \pi_2 f_2(\boldsymbol{z}|w)}{\pi_1 f_1(\boldsymbol{z}|w) + \pi_2 f_2(\boldsymbol{z}|w)}\big]^2 g(x) dw$$

which can be also approximated as the summation form:

$$\sum_{i=1}^{n} \big[\frac{\pi_1 f_1(\boldsymbol{z}|w_i) - \pi_2 f_2(\boldsymbol{z}|w_i)}{\pi_1 f_1(\boldsymbol{z}|w_i) + \pi_2 f_2(\boldsymbol{z}|w_i)}\big]^2$$

[Take off $g(x)$ in summation.]

This is one purity value based on one combination of baseline covariates. Define the purity as $\cancel{P_x}$ $P_{\boldsymbol{\alpha}}$.

We would like to find the $\boldsymbol{\alpha}$ that max the $\cancel{P_x}$ $P_{\boldsymbol{\alpha}}$, i.e.,

$$\hat{\boldsymbol{\alpha}} = \arg\max_{\boldsymbol{\alpha}} P_{\boldsymbol{\alpha}} = \arg\max_{\boldsymbol{\alpha}} \int_{\boldsymbol{\alpha}'\boldsymbol{x}} \left[\frac{\pi_1 f_1(\boldsymbol{z}|\boldsymbol{\alpha}'\boldsymbol{x}) - \pi_2 f_2(\boldsymbol{z}|\boldsymbol{\alpha}'\boldsymbol{x})}{\pi_1 f_1(\boldsymbol{z}|\boldsymbol{\alpha}'\boldsymbol{x}) + \pi_2 f_2(\boldsymbol{z}|\boldsymbol{\alpha}'\boldsymbol{x})}\right]^2 g(\boldsymbol{\alpha}'\boldsymbol{x}) d\boldsymbol{\alpha}'\boldsymbol{x}$$