Outline

Coefficients v.s. Outcome

To calculate the purity based on Kullback-Leibler divergence, we have the following equation:

$$D_{KL}(F_1||F_2) = \int_{-\infty}^{+\infty} f_1(x) \log(\frac{f_1(x)}{f_2(x)})dx \tag{1}$$

where F_1, F_2 are two normal distributions.

For the model:

$$Y = S(\beta + b + \Gamma(\alpha' x)) + \epsilon,$$

we suppose the distribution of the coefficient is from a MVN:

$$Z = \beta + b + \Gamma(\alpha' x) \sim MVN(\beta + \Gamma(\alpha' x), D)$$

where D is the covariance of b.

However, in the EMBARC data, the random effect for the concavity can be relatively small. We may rewrite the model and use the normal distributions of the outcomes to calculate the purity.

$$Y = S(\beta + \Gamma(\alpha' x)) + Ub + \epsilon,$$

And the outcome Y follows

$$Y \sim MVN\big(S(\beta + \Gamma(\alpha' x)), UDU' + \sigma^2 I\big)$$

where $\epsilon \sim N(0, \sigma^2)$.

Prediction

When there is a new patient come to visit, a psychiatrist need to make a personal treatment plan, based on his or her baseline biosignatures, such as age, gender, and some measurements, as well as the treatments effects obtained from previous patients population. While assigned to a treatment group, the patient will be followed for 7 weeks, and his Hamilton Rating Scale for Depression (HRSD) is measured at each follow-up time.

Change score method

To make a decision function, the straightforward idea is fitting a linear regression model, with the outcome difference during the follow-up as the outcome and biosignatures as the independent variables, that is, ignore the longtiudnatla sturcture and use the last - first score as the outcome and the baseline biosignarues serve as independent varables. That is

$$\Delta Y_{i,drg} = Y_{i,7,drg} - Y_{i,0,drg}, \Delta Y_{i,pbo} = Y_{i,7,pbo} - Y_{i,0,pbo}$$

where $\Delta Y_{i,drg}$ is the outcome differences between the first measure and last measure of the ith subject under intervention, while $\Delta Y_{i,pbo}$ is the ith subject's counterfactor outcome differences if he or she takes placebo. If $\Delta Y_{i,drg} > \Delta Y_{i,pbo}$, the patient has a larger decrease in the HRSD when treated with drug, which means that the drug has a better effect than the placebo for that patient and he or she should be assigned into the treatment group. On the other hand, if $\Delta Y_{i,drg} < \Delta Y_{i,pbo}$, the patient has less recovery when treated with drug than had placebo, therefore, the patient does not need drug treatment, since it will not improve the patient's outcome than placebo. By fitting the linear regression

$$\Delta Y_{i,drg} = \gamma_{drg} X, \Delta Y_{i,pbo} = \gamma_{drg} X$$

we can get the estimations of the coefficients $\hat{\gamma}_{drg}, \hat{\gamma}_{drg}$, and the associated estimated outcome difference can be evaluated as:

$$\Delta\hat{Y}_{i,drg} = \hat{\gamma}_{drg}X, \Delta\hat{Y}_{i,pbo} = \hat{\gamma}_{drg}X$$

To estimate the prediction effect of this decision rule, we would like to check whether the esitmated changes of outcome for a subject is consistent with the true changes, which is:

	$\Delta Y_{i,drg} > \Delta Y_{i,pbo}$	$\Delta Y_{i,drg} < \Delta Y_{i,pbo}$
$\hat{\Delta}Y_{i,drg} > \hat{\Delta}Y_{i,pbo}$	1	0
$\hat{\Delta}Y_{i,drg} < \hat{\Delta}Y_{i,pbo}$	0	1

which we defined it as "the concordance of group assignment". The percentage of concordance is used to measure the performance.

Measure	True outcome	Estimated outcome
Fixed effect only	$Y = \beta'X + \epsilon$	$\hat{Y} = \hat{\beta}'X$

Longitudinal average tanget slope approach

After calculate the $\hat{\alpha}$, we may estimate the equation by LME,

$$Y = S(\beta + b + \Gamma(\alpha'x)) + \epsilon,$$

and get the estimated $\hat{\beta}, \hat{\Gamma}, \hat{\alpha}$

The estimated trajectory of the coefficient for drug group can be estimated as

$$\hat{Y}_{drg} = S(\hat{\beta}_{drg} + \hat{\Gamma}_{drg}(\hat{\alpha}'x))$$

The estimated trajectory of placebo group

$$\hat{Y}_{pbo} = S(\hat{\beta}_{pbo} + \hat{\Gamma}_{pbo}(\hat{\alpha}'x))$$

We than compare the average tanget slope, which is also the change score between the end time and start time, given that the intervals have the same length.

- If $\hat{Y}_{drg} > \hat{Y}_{pbo}$, the subject is assigned to placebo group by the estimation
- If $\hat{Y}_{drg} < \hat{Y}_{pbo}$, the subject is assigned to drug group by the estimation

For the true treatment group assignment, we can calculate it from

- Fixed effect only, i.e. $Y = S(\beta + \Gamma(\alpha'x))$
- Fixed + random effect, i.e. $Y = S(\beta + b + \Gamma(\alpha'x))$
- Fixed + random effect + random error, i.e. $Y = S(\beta + b + \Gamma(\alpha'x)) + \epsilon$

To make the sure the comparison is consistent,

Measure	True outcome	Estimated outcome
Fixed effect only	$S\big(\beta + \Gamma(\alpha'x)\big)$	$S\big(\hat{\beta} + \hat{\Gamma}(\hat{\alpha}'x)\big)$
Fixed + random effect	$S\big(\beta + b + \Gamma(\alpha'x)\big)$	$S\big(\hat{\beta} + \hat{\Gamma}(\hat{\alpha}'x)\big)$ / $S\big(\hat{\beta} + \hat{b} + \hat{\Gamma}(\hat{\alpha}'x)\big)$
Fixed + random effect + random error	$S\big(\beta + b + \Gamma(\alpha'x)\big) + \epsilon$	$S\big(\hat{\beta} + \hat{\Gamma}(\hat{\alpha}'x)\big)$