Notes

2020-01-18

In previous results, we fit the outcome variable with a linear mixed model

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

and treat the coefficient $z = \beta + b + \Gamma(\alpha' x)$ as a MVN, that is, $z|w \sim MVN(\beta + \Gamma(\alpha' x), D)$.

We may also calculate the Kullback-Leibler divergence by using the outcome variables directly, by assuming

$$Y = X(\beta + \Gamma(\alpha' x)) + Zb + \epsilon$$
$$Y \sim MVN(S(\beta + \Gamma(\alpha' x)), ZDZ')$$

Let's see whether they can return similar results.

1 Kullback-Leibler divergence and Purity

To measure how much the differences are between the treatment group and the placebo group, we apply the Kullback-Leibler (KL) divergence, which measures how one probability distribution F_1 is different from another probability distribution F_2.

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

where f_1 and f_2 denote the probability density functions (pdf) of F_1 and F_2, separately. The larger the KL divergence between distributions is, the more "pure" the distributions are. Besides, $D_{KL}(F_1||F_2) \geq 0$. Similarly, the $D_{KL}(F_2||F_1)$ is also always larger than or equals to 0.

Based on the Kullback-Leibler divergence, we define the *purity*, which represent how much the differences between the treatment group distribution F_1 and the placebo group distribution F_2. We define the puirty function of the summation of two Kullback-Leibuler divergence as

$$\begin{aligned} purity =& D_{KL}(F_1||F_2) + D_{KL}(F_2||F_1) \\ =& \int_{-\infty}^{+\infty} f_1(x)\log(\frac{f_1(x)}{f_2(x)})dx + \int_{-\infty}^{+\infty} f_2(x)\log(\frac{f_2(x)}{f_1(x)})dx \end{aligned} \tag{2}$$

where

$$f_1(x) \sim MVN(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$$
$$f_2(x) \sim MVN(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$$

Let's calculate the purity value by calculating $\int f_1 \log f_1$, $\int f_2 \log f_2$, $\int f_1 \log f_2$, and $\int f_2 \log f_1$.

Part $\int f_1 \log f_1$

$$\begin{aligned} \int f_1 \log f_1 =& E_1\{-\frac{n}{2}\log(2\pi) - \frac{1}{2}\log(|\boldsymbol{\Sigma}_1|) - \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_1)'(\boldsymbol{\Sigma}_1)^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1)\} \\ =& -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log(|\boldsymbol{\Sigma}_1|) - \frac{1}{2}E_1[(\boldsymbol{x}-\boldsymbol{\mu}_1)'(\boldsymbol{\Sigma}_1)^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1)] \end{aligned}$$

And

$$\begin{aligned}
E_1[(\boldsymbol{x}-\boldsymbol{\mu}_1)'(\boldsymbol{\Sigma}_1)^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1)] =& E_1[tr((\boldsymbol{x}-\boldsymbol{\mu}_1)'(\boldsymbol{\Sigma}_1)^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1))] \\
=& E_1[tr\big((\boldsymbol{\Sigma}_1)^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1)(\boldsymbol{x}-\boldsymbol{\mu}_1)'\big)] \\
=& tr\big(E_1[(\boldsymbol{\Sigma}_1)^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1)(\boldsymbol{x}-\boldsymbol{\mu}_1)']\big) \\
=& tr\big(\Sigma_1^{-1}E_1[(\boldsymbol{x}-\boldsymbol{\mu}_1)(\boldsymbol{x}-\boldsymbol{\mu}_1)']\big) \\
=& tr\big(\Sigma_1^{-1}\Sigma_1\big) = tr(\boldsymbol{I}_n) = n
\end{aligned}$$

Therefore,

$$\int f_1 \log f_1 = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log(|\boldsymbol{\Sigma}_1|) - \frac{n}{2} \tag{3}$$

Similarly,

$$\int f_2 \log f_2 = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log(|\boldsymbol{\Sigma}_2|) - \frac{n}{2} \tag{4}$$

Part $\int f_1 \log f_2$

$$\begin{aligned}
\int f_1 \log f_2 =& E_1(-\frac{n}{2}\log(2\pi) - \frac{1}{2}\log(|\boldsymbol{\Sigma}_2|) - \frac{1}{2}(\boldsymbol{x}-\mu_2)'\Sigma_2^{-1}(\boldsymbol{x}-\mu_2)) \\
=& -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log(|\boldsymbol{\Sigma}_2|) - \frac{1}{2}E_1[(\boldsymbol{x}-\mu_2)'\Sigma_2^{-1}(\boldsymbol{x}-\mu_2)]
\end{aligned}$$

And

$$\begin{aligned}
& E_1[(\boldsymbol{x}-\boldsymbol{\mu}_2)'\boldsymbol{\Sigma}_2^{-1}(x-\boldsymbol{\mu}_2)] \\
=& E_1[(\boldsymbol{x}-\boldsymbol{\mu}_1+\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1+\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)] \\
=& E_1[(\boldsymbol{x}-\boldsymbol{\mu}_1)'\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1) + (\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1) \\
& + (\boldsymbol{x}-\boldsymbol{\mu}_1)\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2) + (\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)] \\
=& E_1[(\boldsymbol{x}-\boldsymbol{\mu}_1)'\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1)] + (\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'\boldsymbol{\Sigma}_2^{-1}E_1(\boldsymbol{x}-\boldsymbol{\mu}_1)+ \\
& E_1(\boldsymbol{x}-\boldsymbol{\mu}_1)')\Sigma_2^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2) + (\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2) \\
=& E_1[(\boldsymbol{x}-\boldsymbol{\mu}_1)'\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1)] + 0 + 0 + (\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2) \\
=& E_1[tr(\boldsymbol{x}-\boldsymbol{\mu}_1)'\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1))] + (\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2) \\
=& E_1[tr(\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1)'(\boldsymbol{x}-\boldsymbol{\mu}_1))] + (\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2) \\
=& tr(E_1[\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1)'(\boldsymbol{x}-\boldsymbol{\mu}_1)]) + (\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2) \\
=& tr(\boldsymbol{\Sigma}_2^{-1}E_1[(x-\boldsymbol{\mu}_1)'(\boldsymbol{x}-\boldsymbol{\mu}_1)]) + (\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2) \\
=& tr(\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}_1) + (\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)
\end{aligned}$$

Therefore,

$$\int f_1 \log f_2 = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log(|\boldsymbol{\Sigma}_2|) - \frac{1}{2}\big\{tr(\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}_1) + (\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)\big\} \tag{5}$$

Similarly,

$$\int f_2 \log f_1 = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log(|\boldsymbol{\Sigma}_1|) - \frac{1}{2}\big\{tr(\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_2) + (\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)\big\} \tag{6}$$

Then the purity is

$$
\begin{aligned}
&\int f_1 \log f_1 + \int f_2 \log f_2 - \int f_2 \log f_1 - \int f_1 \log f_2 \\
=& -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log(|\boldsymbol{\Sigma}_1|) - \frac{n}{2} \\
&-\frac{n}{2}\log(2\pi) - \frac{1}{2}\log(|\boldsymbol{\Sigma}_2|) - \frac{n}{2} \\
&-(-\frac{n}{2}\log(2\pi) - \frac{1}{2}\log(|\boldsymbol{\Sigma}_2|) - \frac{1}{2}\big\{tr(\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}_1) + (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\big\}) \\
&-(-\frac{n}{2}\log(2\pi) - \frac{1}{2}\log(|\boldsymbol{\Sigma}_1|) - \frac{1}{2}\big\{tr(\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_2) + (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\big\}) \\
=& -n + \frac{1}{2}tr(\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_2) + \frac{1}{2}tr(\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}_1) \\
&+\frac{1}{2}[(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)] + \frac{1}{2}[(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)]
\end{aligned} \tag{7}
$$

Then the purity is defined as $-n + \frac{1}{2}tr(\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_2) + \frac{1}{2}tr(\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}_1) + \frac{1}{2}[(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)] + \frac{1}{2}[(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)]$ for two normal distributions f_1, f_2 with mean μ_1, μ_2 respectively and covariance matrix Σ_1, Σ_2 respectively.

Back to our model, when we fit the coefficients of the LME, i,e,

$$z = \boldsymbol{\beta} + \boldsymbol{b} + \boldsymbol{\Gamma}(\boldsymbol{\alpha'x})$$

as multivariate normal distributions and plug in equation (7), we can get our purity function:

$$Purity(\alpha) = A_0 + A_1\boldsymbol{\mu}_x'\boldsymbol{\alpha} + \frac{A_2}{2}\big[\boldsymbol{\alpha}'\boldsymbol{\Sigma}_x\boldsymbol{\alpha} + \boldsymbol{\alpha}'\boldsymbol{\mu}_x\boldsymbol{\mu}_x'\boldsymbol{\alpha}\big] \tag{8}$$

where

$$
\begin{aligned}
A_0 &= -q + \frac{1}{2}tr(\boldsymbol{D}_2^{-1}\boldsymbol{D}_1) + \frac{1}{2}tr(\boldsymbol{D}_1^{-1}\boldsymbol{D}_2) + \frac{1}{2}(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)'(\boldsymbol{D}_1^{-1} + \boldsymbol{D}_2^{-1})(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2) \\
A_1 &= (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)'(\boldsymbol{D}_1^{-1} + \boldsymbol{D}_2^{-1})(\boldsymbol{\Gamma}_1 - \boldsymbol{\Gamma}_2) \\
A_2 &= (\boldsymbol{\Gamma}_1 - \boldsymbol{\Gamma}_2)\big)'(\boldsymbol{D}_1^{-1} + \boldsymbol{D}_2^{-1})\big((\boldsymbol{\Gamma}_1 - \boldsymbol{\Gamma}_2)
\end{aligned}
$$

When we fit the outcome as normal distribution and plug in the equation (7), we can simply replace the $\boldsymbol{\beta}$ in equation (8) as $\boldsymbol{S\beta}$; replace $\boldsymbol{\Gamma}$ as $\boldsymbol{S\Gamma}$; replace $\boldsymbol{D}$ as $\boldsymbol{ZDZ'}$

Replace $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ with $\boldsymbol{X}(\boldsymbol{\beta}_1 + \boldsymbol{\Gamma}_1(\boldsymbol{\alpha'x}))$, and $\boldsymbol{X}(\boldsymbol{\beta}_2 + \boldsymbol{\Gamma}_2(\boldsymbol{\alpha'x}))$. Replace $\boldsymbol{D}_1, \boldsymbol{D}_2$ with $Z\boldsymbol{D}_1\boldsymbol{Z}'$, $Z\boldsymbol{D}_2\boldsymbol{Z}'$. Then the purity function is

$$Purity(\alpha) = B_0 + B_1\boldsymbol{\mu}_x'\boldsymbol{\alpha} + \frac{B_2}{2}\big[\boldsymbol{\alpha}'\boldsymbol{\Sigma}_x\boldsymbol{\alpha} + \boldsymbol{\alpha}'\boldsymbol{\mu}_x\boldsymbol{\mu}_x'\boldsymbol{\alpha}\big] \tag{9}$$

where

$$
\begin{aligned}
B_0 =& -q + \frac{1}{2}tr((Z\boldsymbol{D}_2\boldsymbol{Z}')^{-1}(\boldsymbol{ZD}_1\boldsymbol{Z}')) + \frac{1}{2}tr((\boldsymbol{ZD}_1\boldsymbol{Z}')^{-1}(\boldsymbol{ZD}_2\boldsymbol{Z}')) \\
&+\frac{1}{2}(\boldsymbol{S\beta}_1 - \boldsymbol{S\beta}_2)'((\boldsymbol{ZD}_1\boldsymbol{Z}')^{-1} + (\boldsymbol{ZD}_2\boldsymbol{Z}')^{-1})(\boldsymbol{S\beta}_1 - \boldsymbol{S\beta}_2) \\
B_1 =& (\boldsymbol{S\beta}_1 - \boldsymbol{S\beta}_2)'((\boldsymbol{ZD}_1\boldsymbol{Z}')^{-1} + (\boldsymbol{ZD}_2\boldsymbol{Z}')^{-1})(\boldsymbol{S\Gamma}_1 - \boldsymbol{S\Gamma}_2) \\
B_2 =& (\boldsymbol{S\Gamma}_1 - \boldsymbol{S\Gamma}_2)\big)'((\boldsymbol{ZD}_1\boldsymbol{Z}')^{-1} + (\boldsymbol{ZD}_2\boldsymbol{Z}')^{-1})\big((\boldsymbol{S\Gamma}_1 - \boldsymbol{S\Gamma}_2)
\end{aligned}
$$