Purity Calculation

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

$$f_1(x) \sim MVN(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1), (\mu_1 : p \times 1, \Sigma_1 : p \times p)$$

$$f_2(x) \sim MVN(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2), (\mu_2 : p \times 1, \Sigma_2 : p \times p)$$

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

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

Therefore,

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

Similarly,

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

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,

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

Similarly,

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

Then the purity function is

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

In our settings, we fit the outcome variable with a linear mixed model

$$\boldsymbol{Y}_1 = \boldsymbol{S}(\boldsymbol{\beta}_1 + \boldsymbol{b}_1 + \boldsymbol{\Gamma}_1(\boldsymbol{\alpha}'\boldsymbol{x})) + \boldsymbol{\epsilon}$$

$$\boldsymbol{Y}_2 = \boldsymbol{S}(\boldsymbol{\beta}_2 + \boldsymbol{b}_2 + \boldsymbol{\Gamma}_2(\boldsymbol{\alpha}'\boldsymbol{x})) + \boldsymbol{\epsilon}$$

and treat the coefficient $\boldsymbol{z}_i = \boldsymbol{\beta}_i + \boldsymbol{b}_i + \boldsymbol{\Gamma}_i(\boldsymbol{\alpha}'\boldsymbol{x}), i \in \{1,2\}$ as a MVN, that is,

$$\boldsymbol{z}_1|x \sim MVN(\boldsymbol{\beta}_1 + \boldsymbol{\Gamma}_1(\boldsymbol{\alpha}'\boldsymbol{x}), \boldsymbol{D}_1)$$

$$\boldsymbol{z}_2|x \sim MVN(\boldsymbol{\beta}_2 + \boldsymbol{\Gamma}_2(\boldsymbol{\alpha}'\boldsymbol{x}), \boldsymbol{D}_2)$$

Therefore, $\boldsymbol{\mu}_1 = \boldsymbol{\beta}_1 + \boldsymbol{\Gamma}_1(\boldsymbol{\alpha}'\boldsymbol{x}), \boldsymbol{\mu}_2 = \boldsymbol{\beta}_2 + \boldsymbol{\Gamma}_2(\boldsymbol{\alpha}'x)$, $\boldsymbol{\Sigma}_1 = \boldsymbol{D}_1, \boldsymbol{\Sigma}_2 = \boldsymbol{D}_2$

Therefore

$$\begin{aligned}
&(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'(\boldsymbol{D}_1^{-1} + \boldsymbol{D}_2^{-1})(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \\
&= \big(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2 + (\boldsymbol{\Gamma}_1 - \boldsymbol{\Gamma}_2)\boldsymbol{\alpha}'\boldsymbol{x}\big)'(\boldsymbol{D}_1^{-1} + \boldsymbol{D}_2^{-1})\big(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2 + (\boldsymbol{\Gamma}_1 - \boldsymbol{\Gamma}_2)\boldsymbol{\alpha}'\boldsymbol{x}\big) \\
&= (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)'(\boldsymbol{D}_1^{-1} + \boldsymbol{D}_2^{-1})(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2) \\
&+ 2\big[(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)'(\boldsymbol{D}_1^{-1} + \boldsymbol{D}_2^{-1})(\boldsymbol{\Gamma}_1 - \boldsymbol{\Gamma}_2)\boldsymbol{x}'\boldsymbol{\alpha} \\
&+ \boldsymbol{\alpha}'\boldsymbol{x}\boldsymbol{x}'\boldsymbol{\alpha}\big((\boldsymbol{\Gamma}_1 - \boldsymbol{\Gamma}_2)\big)'(\boldsymbol{D}_1^{-1} + \boldsymbol{D}_2^{-1})\big((\boldsymbol{\Gamma}_1 - \boldsymbol{\Gamma}_2)\big)
\end{aligned}$$

And purity function in terms of $\boldsymbol{\beta}_i, \boldsymbol{\Gamma}_i, \boldsymbol{D}_i$ is,

$$\begin{aligned}
purity = g(\boldsymbol{\alpha}'\boldsymbol{x}) = & -p + \frac{1}{2}tr(\boldsymbol{D}_2^{-1}\boldsymbol{D}_1) + \frac{1}{2}tr(\boldsymbol{D}_1^{-1}\boldsymbol{D}_2) \\
& + \frac{1}{2}\big\{(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)'(\boldsymbol{D}_1^{-1} + \boldsymbol{D}_2^{-1})(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2) \\
& + 2\big[(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)'(\boldsymbol{D}_1^{-1} + \boldsymbol{D}_2^{-1})(\boldsymbol{\Gamma}_1 - \boldsymbol{\Gamma}_2)\boldsymbol{x}'\boldsymbol{\alpha} \\
& + \boldsymbol{\alpha}'\boldsymbol{x}\boldsymbol{x}'\boldsymbol{\alpha}\big((\boldsymbol{\Gamma}_1 - \boldsymbol{\Gamma}_2)\big)'(\boldsymbol{D}_1^{-1} + \boldsymbol{D}_2^{-1})\big((\boldsymbol{\Gamma}_1 - \boldsymbol{\Gamma}_2)\big)\big\}
\end{aligned}$$

The expectation of the purity function is

$$\begin{aligned}
G(\boldsymbol{\alpha}) = E(g(\boldsymbol{\alpha})) = & A_0 + A_1\boldsymbol{\mu}_x'\boldsymbol{\alpha} + \frac{A_2}{2}E[\boldsymbol{\alpha}'\boldsymbol{x}\boldsymbol{x}'\boldsymbol{\alpha}] \\
= & A_0 + A_1\boldsymbol{\mu}_x'\boldsymbol{\alpha} + \frac{A_2}{2}[\boldsymbol{\alpha}'(\boldsymbol{\Sigma}_x + \mu_x'\mu_x)\boldsymbol{\alpha}]
\end{aligned}$$

where

- $E(\boldsymbol{x}) = \boldsymbol{\mu}_x$
- $Var(\boldsymbol{x}) = \boldsymbol{\Sigma}_x$
- $A_0 = -p + \frac{1}{2}tr(\boldsymbol{D}_2^{-1}\boldsymbol{D}_1) + \frac{1}{2}tr(\boldsymbol{D}_1^{-1}\boldsymbol{D}_2) + \frac{1}{2}(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)'(\boldsymbol{D}_1^{-1} + \boldsymbol{D}_2^{-1})(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)$
- $A_1 = (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)'(\boldsymbol{D}_1^{-1} + \boldsymbol{D}_2^{-1})(\boldsymbol{\Gamma}_1 - \boldsymbol{\Gamma}_2)$
- $A_2 = (\boldsymbol{\Gamma}_1 - \boldsymbol{\Gamma}_2)'(\hat{D}_1^{-1} + \boldsymbol{D}_2^{-1})(\boldsymbol{\Gamma}_1 - \boldsymbol{\Gamma}_2)$