STATISTICAL INFERENCE IN ONE-SAMPLE SUMMARY-DATA MENDELIAN RANDOMIZATION

JINGSHU WANG, BAICHEN TAN, YUXUAN KE

Abstract. This is an examination of the methods in the two papers "Powerful three-sample genome-wide design and robust statistical inference in summary-data Mendelian randomization" and "Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score." We propose two adaptations on the previous method: first, we do not use multiple samples; second, we assume the risk factors are multidimensional (i.e. when there exist horizontal pleiotropic effects)

Contents

1. Motivation and problem setup

Notation 1.1. I use γ_j, Γ_j with subscript j to describe each of the jth individual component. and $\gamma = [\gamma_1, \cdots, \gamma_p]^T$, $\Gamma = [\Gamma_1, \cdots, \Gamma_p]^T$ to denote the n dimensional column vector.

Let $\hat{\gamma}_j, \hat{\Gamma}_j, j = 1, 2, \cdots, p$ be the jth snippet's effect on the risk factor and the outcome respectively.

We first write out the log-likelihood for each γ_j

$$l_j(\beta, \gamma) = -\frac{1}{2}\left(\frac{(\hat{\gamma}_j - \gamma_j)^2}{\sigma_{X_j}^2} + \frac{(\hat{\Gamma}_j - \beta\gamma_j)^2}{\sigma_{Y_j}^2}\right)$$

We can solve the MLE estimator for γ_j by taking partial derivative over γ_j

$$\frac{\partial l_j}{\partial \gamma_j} = -\frac{1}{2}\left(\frac{2(\hat{\gamma}_j - \gamma_i)(-1)}{\sigma_{X_j}^2} + \frac{2(\hat{\Gamma}_j - \Gamma_i)(-\beta)}{\sigma_{Y_j}^2}\right) = 0$$

Solving this equation we get

$$\gamma_{j\,\mathrm{MLE}}^* = \frac{W_j(\beta)}{1/\sigma_{X_j}^2 + \beta^2/\sigma_{Y_j}^2} \tag{1.2}$$

where

$$W_j(\beta) = \frac{\hat{\gamma}_j}{\sigma_{X_j}^2} + \frac{\beta\hat{\Gamma}_j}{\sigma_{Y_j}^2} \tag{1.3}$$

We noticed that the function $W_j(\beta)$ is a sufficient statistics for γ_j.

We first notice that after obtaining $W_j(\beta)$, we can therefore write out the form of profile log-likelihood of β as

$$l_j(\beta, \gamma_{j\,\mathrm{MLE}}^*) = -\frac{1}{2}\left(\frac{(\hat{\gamma}_j - \gamma_{j\,\mathrm{MLE}}^*)^2}{\sigma_{X_j}^2} + \frac{(\hat{\Gamma}_j - \beta\gamma_{j\,\mathrm{MLE}}^*)^2}{\sigma_{Y_j}^2}\right)$$

Lemma 1.4. *The profile log-likelihood of* β

$$l_j(\beta, \gamma^*_{j\,\mathrm{MLE}}) = -\frac{1}{2}\left(\frac{(\hat{\gamma}_j - \gamma^*_{j\,\mathrm{MLE}})^2}{\sigma^2_{X_j}} + \frac{(\hat{\Gamma}_j - \beta\gamma^*_{j\,\mathrm{MLE}})^2}{\sigma^2_{Y_j}}\right)$$

can be reduced to a nice form of

$$l_j(\beta) = -\frac{1}{2}\frac{(\hat{\Gamma}_j - \beta\hat{\gamma}_j)^2}{\sigma^2_{X_j}\beta^2 + \sigma^2_{Y_j}} \tag{1.5}$$

Proof. Expanding the profile log-likelihood, we will have

$$-\frac{1}{2}\left(\frac{\hat{\gamma}_j^2 + \gamma^{*2}_{j\,\mathrm{MLE}} - 2\hat{\gamma}_j\gamma^*_{j\,\mathrm{MLE}}}{\sigma^2_{X_j}} + \frac{\hat{\Gamma}_j^2 + \beta^2\gamma^{*2}_{j\,\mathrm{MLE}} - 2\hat{\Gamma}_j\beta\gamma^*_{j\,\mathrm{MLE}}}{\sigma^2_{Y_j}}\right)$$

Reducing the fractions to a common denominator, we have

$$-\frac{1}{2}\left(\frac{\sigma^2_{Y_j}\gamma^{*2}_{j\,\mathrm{MLE}} - 2\sigma^2_{Y_j}\hat{\gamma}_j\gamma^*_{j\,\mathrm{MLE}}}{\sigma^2_{X_j}\sigma^2_{Y_j}} + \frac{\sigma^2_{X_j}\beta^2\gamma^{*2}_{j\,\mathrm{MLE}} - 2\sigma^2_{X_j}\hat{\Gamma}_j\beta\gamma^*_{j\,\mathrm{MLE}}}{\sigma^2_{X_j}\sigma^2_{Y_j}} + \frac{\hat{\gamma}_j^2}{\sigma^2_{X_j}} + \frac{\hat{\Gamma}_j^2}{\sigma^2_{Y_j}}\right)$$

Rearranging to group the coefficients of $\gamma^{*2}_{j\,\mathrm{MLE}}$ and $\gamma^*_{j\,\mathrm{MLE}}$ respectively, we get

$$-\frac{1}{2}\left(\frac{1}{\sigma^2_{X_j}\sigma^2_{Y_j}}\gamma^{*2}_{j\,\mathrm{MLE}}\left(\sigma^2_{Y_j} + \sigma^2_{X_j}\beta^2\right) - \frac{1}{\sigma^2_{X_j}\sigma^2_{Y_j}}2\left(\hat{\gamma}_j\sigma^2_{Y_j} + \hat{\Gamma}_j\beta\sigma^2_{X_j}\right)\gamma^*_{j\,\mathrm{MLE}} + \frac{\hat{\gamma}_j^2}{\sigma^2_{X_j}} + \frac{\hat{\Gamma}_j^2}{\sigma^2_{Y_j}}\right) \tag{1.6}$$

We first focus on the first term, expanding $\gamma^*_{j\,\mathrm{MLE}}$ by equation 1.2, we have

$$\begin{aligned}
&\frac{1}{\sigma^2_{X_j}\sigma^2_{Y_j}} \times \frac{W_j(\beta)^2}{\frac{1}{\sigma^4_{X_j}} + \frac{\beta^4}{\sigma^4_{Y_j}} + \frac{2\beta^2}{\sigma^2_{X_j}\sigma^2_{Y_j}}} \times (\sigma^2_{Y_j} + \sigma^2_{X_j}\beta^2) \\
&= \frac{W_j(\beta)^2}{\frac{\sigma^2_{Y_j}}{\sigma^2_{X_j}} + \frac{\beta^4\sigma^2_{X_j}}{\sigma^2_{Y_j}} + 2\beta^2}(\sigma^2_{Y_j} + \sigma^2_{X_j}\beta^2) \\
&= \frac{W_j(\beta)^2}{\frac{\sigma^4_{Y_j} + \beta^4\sigma^4_{X_j} + 2\beta^2\sigma^2_{X_j}\sigma^2_{Y_j}}{\sigma^2_{X_j}\sigma^2_{Y_j}}}(\sigma^2_{Y_j} + \sigma^2_{X_j}\beta^2) \\
&= \frac{W_j(\beta)^2}{(\sigma^2_{Y_j} + \sigma^2_{X_j}\beta^2)^2}(\sigma^2_{Y_j} + \sigma^2_{X_j}\beta^2)\sigma^2_{X_j}\sigma^2_{Y_j} \\
&= \frac{W_j(\beta)^2}{\sigma^2_{Y_j} + \sigma^2_{X_j}\beta^2}\sigma^2_{X_j}\sigma^2_{Y_j} \\
&= \frac{1}{\sigma^2_{Y_j} + \sigma^2_{X_j}\beta^2}\left(\frac{\hat{\gamma}_j^2}{\sigma^4_{X_j}} + \frac{\beta^2\hat{\Gamma}^2}{\sigma^4_{Y_j}} + \frac{2\hat{\gamma}_j\hat{\Gamma}_j\beta}{\sigma^2_{X_j}\sigma^2_{Y_j}}\right)\sigma^2_{X_j}\sigma^2_{Y_j} \\
&= \frac{1}{\sigma^2_{Y_j} + \sigma^2_{X_j}\beta^2}\left(\frac{\hat{\gamma}_j^2\sigma^2_{Y_j}}{\sigma^2_{X_j}} + \frac{\beta^2\hat{\Gamma}^2\sigma^2_{X_j}}{\sigma^2_{Y_j}} + 2\hat{\gamma}_j\hat{\Gamma}_j\beta\right)
\end{aligned}$$

We then expand the second term of equation 1.6.

$$
\begin{aligned}
&- \frac{1}{\sigma_{X_j}^2 \sigma_{Y_j}^2} 2\left(\hat{\gamma}_j \sigma_{Y_j}^2 + \hat{\Gamma}_j \beta \sigma_{X_j}^2\right) \gamma_{j\,\mathrm{MLE}}^{*} \\
&= -\frac{1}{\sigma_{X_j}^2 \sigma_{Y_j}^2} \left(\frac{W_j(\beta)}{\frac{\sigma_{Y_j}^2 + \sigma_{X_j}^2 \beta^2}{\sigma_{X_j}^2 \sigma_{Y_j}^2}} \right) 2\left(\hat{\gamma}_j \sigma_{Y_j}^2 + \hat{\Gamma}_j \beta \sigma_{X_j}^2\right) \\
&= -\frac{1}{\sigma_{Y_j}^2 + \sigma_{X_j}^2 \beta^2} 2W_j(\beta) \left(\hat{\gamma}_j \sigma_{Y_j}^2 + \hat{\Gamma}_j \beta \sigma_{X_j}^2\right) \\
&= -\frac{1}{\sigma_{Y_j}^2 + \sigma_{X_j}^2 \beta^2} \left(\frac{\hat{\gamma}_j}{\sigma_{X_j}^2} + \frac{\beta \hat{\Gamma}_j}{\sigma_{Y_j}^2} \right) \left(2\hat{\gamma}_j \sigma_{Y_j}^2 + 2\hat{\Gamma}_j \beta \sigma_{X_j}^2\right) \\
&= -\frac{1}{\sigma_{Y_j}^2 + \sigma_{X_j}^2 \beta^2} \left(\frac{2\hat{\gamma}_j^2 \sigma_{Y_j}^2}{\sigma_{X_j}^2} + 2\hat{\gamma}_j \hat{\Gamma}_j \beta + 2\beta \hat{\Gamma}_j \hat{\gamma}_j + \frac{2\beta^2 \hat{\Gamma}_j^2 \sigma_{X_j}^2}{\sigma_{Y_j}^2} \right) \\
&= -\frac{1}{\sigma_{Y_j}^2 + \sigma_{X_j}^2 \beta^2} \left(\frac{2\hat{\gamma}_j^2 \sigma_{Y_j}^2}{\sigma_{X_j}^2} + 4\hat{\gamma}_j \hat{\Gamma}_j \beta + \frac{2\beta^2 \hat{\Gamma}_j^2 \sigma_{X_j}^2}{\sigma_{Y_j}^2} \right)
\end{aligned}
$$

For the last two terms, we can see that

$$
\begin{aligned}
&\frac{\hat{\gamma}_j^2}{\sigma_{X_j}^2} + \frac{\hat{\Gamma}_j^2}{\sigma_{Y_j}^2} \\
&= \frac{1}{\sigma_{Y_j}^2 + \sigma_{X_j}^2 \beta^2} \left(\frac{\hat{\gamma}_j^2}{\sigma_{X_j}^2} (\sigma_{Y_j}^2 + \sigma_{X_j}^2 \beta^2) + \frac{\hat{\Gamma}_j^2}{\sigma_{Y_j}^2} (\sigma_{Y_j}^2 + \sigma_{X_j}^2 \beta^2) \right) \\
&= \frac{1}{\sigma_{Y_j}^2 + \sigma_{X_j}^2 \beta^2} \left(\frac{\hat{\gamma}_j^2 \sigma_{Y_j}^2}{\sigma_{X_j}^2} + \hat{\gamma}_j^2 \beta^2 + \hat{\Gamma}_j^2 + \frac{\hat{\Gamma}_j^2 \beta^2 \sigma_{X_j}^2}{\sigma_{Y_j}^2} \right)
\end{aligned}
$$

Then combining the three parts, we will write out the full form of equation 1.6 as

$$
\begin{aligned}
&-\frac{1}{2} \frac{1}{\sigma_{Y_j}^2 + \sigma_{X_j}^2 \beta^2} \left(\frac{\hat{\gamma}_j^2 \sigma_{Y_j}^2}{\sigma_{X_j}^2} + \frac{\beta^2 \hat{\Gamma}^2 \sigma_{X_j}^2}{\sigma_{Y_j}^2} + 2\hat{\gamma}_j \hat{\Gamma}_j \beta - \frac{2\hat{\gamma}_j^2 \sigma_{Y_j}^2}{\sigma_{X_j}^2} - 4\hat{\gamma}_j \hat{\Gamma}_j \beta - \frac{2\beta^2 \hat{\Gamma}_j^2 \sigma_{X_j}^2}{\sigma_{Y_j}^2} + \frac{\hat{\gamma}_j^2 \sigma_{Y_j}^2}{\sigma_{X_j}^2} + \hat{\gamma}_j^2 \beta^2 + \hat{\Gamma}_j^2 + \frac{\hat{\Gamma}_j^2 \beta^2 \sigma_{X_j}^2}{\sigma_{Y_j}^2} \right) \\
&= -\frac{1}{2} \frac{1}{\sigma_{Y_j}^2 + \sigma_{X_j}^2 \beta^2} \left(\hat{\gamma}_j^2 \beta^2 + \hat{\Gamma}_j^2 - 2\beta \hat{\gamma}_j \hat{\Gamma}_j \right) = -\frac{1}{2} \frac{(\hat{\Gamma}_j - \beta \hat{\gamma}_j)^2}{\sigma_{Y_j}^2 + \sigma_{X_j}^2 \beta^2}
\end{aligned}
$$

□

To solve for the optimal beta, we will then take the derivative of the profile log-likelihood over β. We define the profile score to be the derivative of the profile log-likelihood of β

Definition 1.7. The profile score function $\psi(\beta)$ is the derivative of the profile log-likelihood of β.

$$
\psi(\beta) = -l'(\beta) = \frac{(\hat{\Gamma}_j - \beta \hat{\gamma}_j)(\hat{\Gamma}_j \sigma_{X_j}^2 \beta + \hat{\gamma}_j \sigma_{Y_j}^2)}{(\sigma_{X_j}^2 \beta^2 + \sigma_{Y_j}^2)^2}
$$

We find the optimal $\hat{\beta}_{\mathrm{MLE}}^{*}$ by setting $\psi(\beta) = 0$. Because the MLE solution $\hat{\beta}_{\mathrm{MLE}}^{*}$ is consistent, it will asymptotically converges to the true β_0. The MLE method is valid because at true beta β_0, $\mathbb{E}[\psi(\beta_0)]$ will indeed be equal to 0. To show why this is true, we first prove the following lemma.

Lemma 1.8. *The profile log-likelihod of β in Lemma 1.4 can be written as*

$$
l(\beta) = -\frac{1}{2} \frac{\gamma_j^2 (\beta_0 - \beta)^2 + (e_j - \beta \epsilon_j)^2 + 2\gamma_j (\beta_0 - \beta)(e_j - \beta \epsilon_j)}{\sigma_{X_j}^2 \beta^2 + \sigma_{Y_j}^2}
$$

where $e_j = \hat{\Gamma}_j - \Gamma_j$ and $\epsilon_j = \hat{\gamma}_j - \gamma_j$.

Proof. It is sufficient to prove that the numerator of the two forms matches with each other. We expand the numerator in Lemma 1.8 and obtain

$$\gamma_j^2\beta_0^2 + \gamma_j^2\beta^2 - 2\gamma_j^2\beta_0\beta + e_j^2 + \beta^2\epsilon_j^2 - 2e_j\epsilon_j\beta + 2\gamma_j\beta_0 e_j - 2\gamma_j\beta_0\beta\epsilon_j - 2\gamma_j\beta e_j + 2\gamma_j\beta^2\epsilon_j$$

Substituting $e_j = \hat{\Gamma}_j - \Gamma_j$ and $\epsilon_j = \hat{\gamma}_j - \gamma_j$, we will obtain

$$\begin{aligned}&\gamma_j^2\beta_0^2 + \gamma_j^2\beta^2 - 2\gamma_j^2\beta_0\beta + (\hat{\Gamma}_j - \Gamma_j)^2 + \beta^2(\hat{\gamma}_j - \gamma_j)^2 - 2(\hat{\Gamma}_j - \Gamma_j)(\hat{\gamma}_j - \gamma_j)\beta + 2\gamma_j\beta_0(\hat{\Gamma}_j - \Gamma_j)\\ &- 2\gamma_j\beta_0\beta(\hat{\gamma}_j - \gamma_j) - 2\gamma_j\beta(\hat{\Gamma}_j - \Gamma_j) + 2\gamma_j\beta^2(\hat{\gamma}_j - \gamma_j)\end{aligned}$$

Using the trick that $\Gamma_j = \beta_0\gamma_j$ and canceling out the terms, we will get

$$\hat{\Gamma}_j^2 + \beta^2\hat{\gamma}_j^2 - 2\hat{\Gamma}_j\hat{\gamma}_j = (\hat{\Gamma}_j - \beta\hat{\gamma}_j)^2$$

□

Theorem 1.9. *The expectation of the profile score function ψ will have mean 0 at the true beta β_0. Namely,*

$$\mathbb{E}[\psi(\beta_0)] = 0$$

Proof. We first notice that by Lemma 1.8 the profile score function in definition 1.7 can be rewritten as

$$\psi(\beta_0) = \frac{(e_j - \beta_0\epsilon_j)[\gamma_j(\sigma_{Y_j}^2 + \sigma_{X_j}^2\beta_0^2) + e_j\sigma_{X_j}^2\beta_0 + \epsilon_j\sigma_{Y_j}^2]}{(\sigma_{Y_j}^2 + \sigma_{X_j}^2\beta_0^2)^2}$$

Taking expectation over this, we will have $\mathbb{E}[e_j] = 0$ and $\mathbb{E}[\epsilon_j] = 0$. We also have $\mathrm{Cov}(e_j, \epsilon_j) = 0$ because they are independent. Hence, the expectation $\mathbb{E}[\psi(\beta_0)]$ can be simplified as

$$\mathbb{E}[\psi(\beta_0)] = \frac{\mathbb{E}[e_j^2\sigma_{X_j}^2\beta_0 - \epsilon_j^2\sigma_{Y_j}^2\beta_0]}{(\sigma_{Y_j}^2 + \sigma_{Y_j}^2\beta_0^2)^2} \tag{1.10}$$

Notice that the expectation of $\mathbb{E}[e_j^2]$ is equal to

$$\begin{aligned}\mathbb{E}[e_j^2] &= \mathbb{E}[(\hat{\Gamma}_j - \Gamma_j)^2]\\ &= \mathbb{E}[\hat{\Gamma}_j^2 + \Gamma_j^2 - 2\Gamma_j\hat{\Gamma}_j]\\ &= \mathbb{E}[\hat{\Gamma}_j^2] + \mathbb{E}[\Gamma_j^2] - 2\Gamma_j\mathbb{E}[\hat{\Gamma}_j]\\ &= \sigma_{Y_j}^2 + \Gamma_j^2 + \Gamma_j^2 - 2\Gamma_j^2\\ &= \sigma_{Y_j}^2\end{aligned}$$

where the second last equation holds because $\mathbb{E}[\hat{\Gamma}_j^2] = \mathrm{Var}(\hat{\Gamma}_j^2) + \mathbb{E}[\hat{\Gamma}_j]^2 = \sigma_{Y_j}^2 + \Gamma_j^2$ and $\mathbb{E}[\hat{\Gamma}_j] = \Gamma_j$. By a similar approach we can show that $\mathbb{E}[\epsilon_j^2] = \sigma_{X_j}^2$ and therefore the denominator in formula 1.10 is equal to

$$\mathbb{E}[e_j^2\sigma_{X_j}^2\beta_0 - \epsilon_j^2\sigma_{Y_j}^2\beta_0] = \mathbb{E}[e_j^2]\sigma_{X_j}^2\beta_0 - \mathbb{E}[\epsilon_j^2]\sigma_{Y_j}^2\beta_0 = \sigma_{Y_j}^2\sigma_{X_j}^2\beta_0 - \sigma_{X_j}^2\sigma_{Y_j}^2\beta_0 = 0$$

□

We can further show the consistency of $\hat{\beta}^*_{\mathrm{MLE}}$ by showing that $\hat{\beta}^*_{\mathrm{MLE}}$ converges to β_0 in probability and $\hat{\beta}^*_{\mathrm{MLE}} - \beta_0$ converges to normal distribution. The proof of these two theorems are fully discussed in the paper [1]. Here we state the main theorem.

Theorem 1.11. *Under the assumptions specified in the paper [1], $\hat{\beta}^*_{\mathrm{MLE}} - \beta_0$ converges to normal distribution. Specifically*

$$\frac{V_2}{\sqrt{V_1}}(\hat{\beta}^*_{\mathrm{MLE}} - \beta_0) \xrightarrow{d} N(0,1)$$

where

$$V_1 = \sum_{j=1}^{p} \frac{\gamma_j^2\sigma_{Y_j}^2 + \Gamma_j^2\sigma_{X_j}^2 + \sigma_{X_j}^2\sigma_{Y_j}^2}{(\sigma_{X_j}^2\beta_0^2 + \sigma_{Y_j}^2)} \qquad V_2 = \sum_{j=1}^{p} \frac{\gamma_j^2\sigma_{Y_j}^2 + \sigma_{X_j}^2\sigma_{Y_j}^2}{(\sigma_{X_j}^2\beta_0^2 + \sigma_{Y_j}^2)}$$

Proof. The proof of this theorem is fully discussed in the paper [1] □

It is important to note that in the above setting, weak indicator variables will decrease the efficiency. Suppose we have a new snippet Z_{p+1} that is independent of the risk fadtor X. Then $\gamma_{p+1} = 0$ and so V_1 increases but V_2 stays the same, resulting a larger variance of $\hat{\beta}^*_{\text{MLE}}$.

The profile-likelihood method holds under the assumption that there is no pleiotropic effect, i.e., $\Gamma_j = \beta_0 \gamma_j$. To show that the MLE estimator $\hat{\beta}^*_{\text{MLE}}$ is not consistent when there is a pleiotropic effect, we can assume that $\alpha_j = \Gamma_j - \beta_0\gamma_j \sim N(0, \tau_0^2)$, where α_j measures the pleiotropic effect of the jth snippet and τ_0^2 is some unknown additive constant that inflates the variance of $\hat{\Gamma}_j$. Then we have the following setup:

$$\hat{\gamma}_j \sim N(\gamma_j, \sigma^2_{X_j}) \qquad \hat{\Gamma}_j \sim N(\gamma_j\beta_0, \sigma^2_{Y_j} + \tau_0^2),\ j \in [p]$$

Then the corresponding profile likelihood is

$$l(\beta, \tau^2) = -\frac{1}{2}\sum_{j=1}^{p} \frac{(\hat{\Gamma}_j - \beta\hat{\gamma}_j)^2}{\sigma^2_{X_j}\beta^2 + \sigma^2_{Y_j} + \tau^2} + \log(\sigma^2_{Y_j} + \tau^2)$$

In the profile likelihood method, we will take derivative of $l(]\beta, \tau^2)$ over β and τ and solve the equation

$$\frac{\partial}{\partial\beta} l(\beta, \tau^2) = 0 \qquad \frac{\partial}{\partial\tau^2} l(\beta, \tau^2) = 0$$

While the first estimating equation is unbiased such that its expectation is equal to 0 at true (β_0, τ_0^2), the other profile score

$$\frac{\partial}{\partial\tau^2} l(\beta, \tau^2) = \frac{1}{2}\sum_{j=1}^{p} \frac{(\hat{\Gamma}_j - \beta\hat{\gamma}_j)^2}{(\sigma^2_{X_j}\beta^2 + \sigma^2_{Y_j} + \tau^2)^2} - \frac{1}{\sigma^2_{Y_j} + \tau^2}$$

which is not equal to 0 at the true parameters (β_0, τ_0^2). Therefore the MLE estimator when there is pleiotropic effect is not consistent.

2. Statistical Methods

To deal with the two issues induced by the profile likelihood method, we first revisit the formula for the profile score function defined in Definition 1.7 and consider the simple case when there is no pleiotropic effect.

$$\begin{aligned}
\psi(\beta) &= \frac{(\hat{\Gamma}_j - \beta\hat{\gamma}_j)(\hat{\Gamma}_j\sigma^2_{X_j}\beta + \hat{\gamma}_j\sigma^2_{Y_j})}{(\sigma^2_{X_j}\beta^2 + \sigma^2_{Y_j})^2} \\
&= \frac{(\hat{\Gamma}_j - \beta\hat{\gamma}_j)}{\sigma^2_{X_j}\beta^2 + \sigma^2_{Y_j}} \cdot \frac{\hat{\Gamma}_j\sigma^2_{X_j}\beta + \hat{\gamma}_j\sigma^2_{Y_j}}{\sigma^2_{X_j}\beta^2 + \sigma^2_{Y_j}} \\
&= \frac{(\hat{\Gamma}_j - \beta\hat{\gamma}_j)}{\sigma^2_{X_j}\beta^2 + \sigma^2_{Y_j}} \cdot \gamma^*_{j\,\text{MLE}}
\end{aligned}$$

where the last equation holds because

$$\frac{\hat{\Gamma}_j\sigma^2_{X_j}\beta + \hat{\gamma}_j\sigma^2_{Y_j}}{\sigma^2_{X_j}\beta^2 + \sigma^2_{Y_j}} = \frac{(\hat{\Gamma}_j\sigma^2_{X_j}\beta + \hat{\gamma}_j\sigma^2_{Y_j})\frac{1}{\sigma^2_{X_j}\sigma^2_{Y_j}}}{(\sigma^2_{X_j}\beta^2 + \sigma^2_{Y_j})\frac{1}{\sigma^2_{X_j}\sigma^2_{Y_j}}} = \gamma^*_{j\,\text{MLE}}$$

Then we can see that the profile score function has a nice form of

$$\frac{(\hat{\Gamma}_j - \beta\hat{\gamma}_j)}{\sigma^2_{X_j}\beta^2 + \sigma^2_{Y_j}} \cdot \gamma^*_{j\,\text{MLE}} \tag{2.1}$$

We first observe that the term $\hat{\Gamma}_j - \beta\hat{\gamma}_j$ is the "regression residual" of regressing $\hat{\Gamma}_j \sim \hat{\gamma}_j$. At the true β_0, the mean of the residual is equal to 0.

Lemma 2.2. *The mean of the residual $\hat{\Gamma}_j - \beta\hat{\gamma}_j$ is equal to 0 at the true beta β_0. Namely,*

$$\mathbb{E}[\hat{\Gamma}_j - \beta_0\hat{\gamma}_j] = 0$$

Proof. By the linearity of expectation

$$\mathbb{E}[\hat{\Gamma}_j - \beta_0\hat{\gamma}_j] = \mathbb{E}[\hat{\Gamma}_j] - \beta_0\mathbb{E}[\hat{\gamma}_j] = \Gamma_j - \beta_0\gamma_j = 0$$

□

Second, we can see that in formula 2.1, $\gamma^*_{j\,\mathrm{MLE}}$ acts only as a weight to the regression residual $\hat{\Gamma}_j - \beta\hat{\gamma}_j$.
Third, the sufficient statistics $W_j(\beta)$ of γ_j is independent of the regression residual $\hat{\Gamma}_j - \beta\hat{\gamma}_j$.

Lemma 2.3. *The regression residual $\hat{\Gamma}_j - \beta\hat{\gamma}_j$ is independent of $W_j(\beta)$*

Proof. We proved this by showing that their covariance is 0 because the residual and $W_j(\beta)$ are both Gaussian variables.

$$\begin{aligned}\mathrm{Cov}(\hat{\Gamma}_j - \beta\hat{\gamma}_j, W_j(\beta)) &= \mathrm{Cov}(\hat{\Gamma}_j - \beta\hat{\gamma}_j, \frac{\hat{\gamma}_j}{\sigma^2_{X_j}} + \frac{\beta\hat{\Gamma}_j}{\sigma^2_{Y_j}})\\ &= \mathrm{Cov}(\hat{\Gamma}_j, \frac{\hat{\gamma}_j}{\sigma^2_{X_j}}) + \mathrm{Cov}(\hat{\Gamma}_j, \frac{\beta\hat{\Gamma}_j}{\sigma^2_{Y_j}}) - \mathrm{Cov}(\beta\hat{\gamma}_j, \frac{\hat{\gamma}_j}{\sigma^2_{X_j}}) - \mathrm{Cov}(\beta\hat{\gamma}_j, \frac{\beta\hat{\Gamma}_j}{\sigma^2_{Y_j}})\end{aligned}$$

We have $\mathrm{Cov}(\hat{\Gamma}_j, \hat{\gamma}_j) = 0$ because we assume that they are mutually independent. Therefore, the covariance $\mathrm{Cov}(\hat{\Gamma}_j - \beta\hat{\gamma}_j, W_j(\beta))$ is equal to

$$\mathrm{Var}(\hat{\Gamma}_j)\frac{\beta}{\sigma^2_{Y_j}} - \mathrm{Var}(\hat{\gamma}_j)\frac{\beta}{\sigma^2_{X_j}} = \sigma^2_{Y_j}\frac{\beta}{\sigma^2_{Y_j}} - \sigma^2_{X_j}\frac{\beta}{\sigma^2_{X_j}} = 0$$

□

Therefore, to sum up the three observations we made in the formula 2.1.

- The mean of the residual $\hat{\Gamma}_j - \beta\hat{\gamma}_j$ is equal to 0 at the true beta β_0. $\mathbb{E}[\hat{\Gamma}_j - \beta_0\hat{\gamma}_j] = 0$
- $\gamma^*_{j\,\mathrm{MLE}}$ acts only as a weight to the regression residual $\hat{\Gamma}_j - \beta\hat{\gamma}_j$
- The regression residual $\hat{\Gamma}_j - \beta\hat{\gamma}_j$ is independent of $W_j(\beta)$

Recall that $\gamma^*_{j\,\mathrm{MLE}}$ can be thought as a function of $W_j(\beta)$. Then intuitively, we can also plug in any estimator γ^*_j of γ_j (not necessarily the MLE estimator $\gamma^*_{j\,\mathrm{MLE}}$) that only depends on the the sufficient statistics $W_j(\beta)$. We can therefore define a new "score" function $C_j(\beta, \gamma^*_j)$ that is similar to the profile score function in Definition 1.7, which helps find the optimal β^* by solving the equation $C_j(\beta, \gamma^*_j) = 0$.

Definition 2.4. The conditional score function is defined as

$$C(\beta, \gamma^*_j) = \frac{(\hat{\Gamma}_j - \beta\hat{\gamma}_j)}{\sigma^2_{X_j}\beta^2 + \sigma^2_{Y_j}} \cdot \gamma^*_j$$

where $\gamma^*_j = f(W_j(\beta), \beta)$ is an estimator of γ_j through the function $f(W_j(\beta), \beta)$.

For now we are only focusing on one sinppet, and aggregating all the p snippets, we can define the total conditional score function as

Definition 2.5. The total conditional score function it defined as the sum of all p snippets' conditional score function

$$C(\beta) = \sum_{j=1}^{p} C_j(\beta, \gamma^*_j) = \sum_{j=1}^{p} \frac{(\hat{\Gamma}_j - \beta\hat{\gamma}_j)}{\sigma^2_{X_j}\beta^2 + \sigma^2_{Y_j}} \cdot \gamma^*_j$$

where $\gamma^*_j = f(W_j(\beta), \beta)$ is an estimator of γ_j through the function $f(W_j(\beta), \beta)$.

We then solve the optimal β^* by solving the equation $C(\beta) = 0$. The obtained estimator β^* is a reasonable estimator of the true beta β_0 because $\mathbb{E}[C(\beta_0)] = 0$

Lemma 2.6. *The mean of $C(\beta)$ is equal to 0 at the true beta β_0. Namely,*

$$\mathbb{E}[C(\beta_0)] = 0$$

Proof. We first notice that γ_j^* is independent of the residual $\hat{\Gamma}_j - \beta\hat{\gamma}_j$ because $W_j(\beta)$ is independent of the residual $\hat{\Gamma}_j - \beta\hat{\gamma}_j$. Therefore, for each $C_j(\beta, \gamma_j^*)$,

$$\mathbb{E}[C_j(\beta, \gamma_j^*)] = \frac{1}{\sigma_{X_j}^2\beta^2 + \sigma_{Y_j}^2}\mathbb{E}[\hat{\Gamma}_j - \beta\hat{\gamma}_j]\mathbb{E}[\gamma_j^*]$$

We have shown that $\mathbb{E}[\hat{\Gamma}_j - \beta\hat{\gamma}_j] = 0$ at the true beta β_0 and therefore the overall expectation is equal to 0 □

To generalize the idea in Definition 2.5, we construct the following generalized score function

Definition 2.7. The generalized estimating function of β is defined as

$$C(\beta) = \sum_{j=1}^{p} \frac{f_j(W_j(\beta), \beta)\phi(t_j(\beta))}{\sqrt{\beta^2\sigma_{X_j}^2 + \sigma_{Y_j}^2}}$$

where $t_j(\beta) = \frac{\hat{\Gamma}_j - \beta\hat{\gamma}_j}{\sqrt{\beta^2\sigma_{X_j}^2 + \sigma_{Y_j}^2}}$ and function ϕ is some odd function $\phi(-t_j) = -\phi(t_j)$ (Why?)

Remark 2.8. We need the function ϕ to be odd because

We have already proved that under the non-pleiotropic effect assumption, solving $C(\beta)$ will always give us an unbiased estimator. However, under the existence of weak indicator variables, the statistical efficiency is really low. We have previously mentioned that under the existence of weak indicator variable, the variance of the MLE estimator $\hat{\beta}^*_{\text{MLE}}$ is really large. To examine this issue with another perspective, assume that most γ_j are equal to 0. Then intuitively the jth snippet provides no information on β because the distribution of $(\hat{\gamma}_j, \hat{\Gamma}_j)$ is independent of β. However, as we can see in the form of MLE estimator $\gamma^*_{j\,\text{MLE}}$, we are still using some information to estimate γ_j, resulting in a loss of statistical efficiency. Therefore, ideally, to solve the issue of the existence of weak indicator variables, we aim to seek for other more powerful estimator of γ_j^*.

3. An empirical bayes approach

In this section, we propose to estimate γ_j^* by the empirical Bayes estimator and want to show that this improves the statistical efficiency. Assume that the effect β is fixed and also assume that

$$\hat{\gamma}_j \mid \gamma_j \overset{\text{i.i.d.}}{\sim} N(\gamma_j, \sigma_{X_j}^2) \qquad \hat{\Gamma}_j \mid \gamma_j \overset{\text{i.i.d.}}{\sim} N(\gamma_j\beta, \sigma_{Y_j}^2)$$

We also assume that the prior distribution of γ is modeled by a parametric family $\pi(\eta)$ for some parameter η.

$$\gamma_j \mid \eta \overset{\text{i.i.d.}}{\sim} \pi(\eta)$$

Then the empirical Bayes approach will states the follow: to estimate the posterior distribution of $p(\gamma_j \mid \hat{\gamma}_j, \hat{\Gamma}_j, \beta, \eta)$, we can expand the Bayes rule to see that

$$\begin{aligned} p(\gamma_j \mid W_j(\beta), \beta, \eta) &= \frac{p(\gamma_j \mid \eta, \beta)p(W_j(\beta) \mid \gamma_j, \beta, \eta)}{p(W_j(\beta) \mid \eta, \beta)} \\ &= \frac{\pi(\eta)p(W_j(\beta) \mid \gamma_j, \beta, \eta)}{p(W_j(\beta) \mid \eta, \beta)} \end{aligned}$$

where the last equation holds because the distribution of γ_j doesn't depend on β.

Then let $\hat{\eta}$ be an estimator of η based on the function $p(\hat{\gamma}, \hat{\Gamma}_j \mid \eta)$ (for example, by MLE method). We can then calculate the posterior $p(\gamma_j \mid W_j(\beta), \eta)$ and we estimate the γ_j^* to be the mean of the posterior distribution. Then we have

$$\hat{\gamma}^*_{j\,\text{EB}} = \mathbb{E}_{\pi(\hat{\eta})}[\gamma_j \mid W_j(\beta)] \tag{3.1}$$

We know introduce the details of the implementation. We choose the Gaussian mixture distribution as our parametric family for γ_j. Namely, we assume

Model 1. *We assume the prior distribution of γ_j is a Gaussian mixture distribution*

$$\hat{\gamma}_j/\sigma_{X_j}^2 \overset{i.i.d.}{\sim} p_1 \cdot N(0, \sigma_1^2 + 1) + (1 - p_1) \cdot N(0, \sigma_2^2 + 1)$$

for $j = 1, 2, \cdots, p$

We can estimate the posterior mean of γ_j/σ_{X_j} by the following proposition

Proposition 3.2. *Proposition 1. Suppose $Z \sim \mathrm{N}\left(\gamma, \sigma^2\right), \gamma \sim p_1\ \mathrm{N}\left(\mu_1, \sigma_1^2\right) + (1 - p_1)\,\mathrm{N}\left(\mu_2, \sigma_2^2\right)$, then $\gamma \mid Z \sim \tilde{p} \cdot \mathrm{N}\left(\tilde{\mu}_1, \tilde{\sigma}_1^2\right) + (1 - \tilde{p}) \cdot \mathrm{N}\left(\tilde{\mu}_2, \tilde{\sigma}_2^2\right)$, where*

$$\tilde{\mu}_k = \frac{Z/\sigma^2 + \mu_k/\sigma_k^2}{1/\sigma^2 + 1/\sigma_k^2}, \tilde{\sigma}_k^2 = \frac{1}{1/\sigma^2 + 1/\sigma_k^2}, \text{ and}$$

$$\tilde{p} = \frac{p_1 \cdot \varphi\left(Z; \mu_1, \sigma^2 + \sigma_1^2\right)}{p_1 \cdot \varphi\left(Z; \mu_1, \sigma^2 + \sigma_1^2\right) + (1 - p_1) \cdot \varphi\left(Z; \mu_2, \sigma^2 + \sigma_2^2\right)}.$$

In the above equation, $\varphi\left(z; \mu, \sigma^2\right)$ is the probability density function of the normal distribution $\mathrm{N}\left(\mu, \sigma^2\right)$: $\varphi\left(z; \mu, \sigma^2\right) = \left(\sqrt{2\pi\sigma^2}\right)^{-1} \exp\left\{-(z - \mu)^2/\left(2\sigma^2\right)\right\}$. The posterior mean of γ is given by $\hat{\gamma} = \mathbb{E}[\gamma \mid Z] = \tilde{p}\tilde{\mu}_1 + (1 - \tilde{p})\tilde{\mu}_2$.

Proof. We define $I \sim \text{Bernoulli}(p_1)$ such that $\Pr(I = 0) = p_1$ and $\Pr(I = 1) = 1 - p_1$. We use $p_0(i)$ to denote the probability mass function of the variable I. Notice that the gaussian mixture posterior can be written as

$$p(\gamma \mid Z) = \sum_{i=0}^{1} p(I = i \mid Z) \cdot p(\gamma \mid Z, I = i)$$

where

$$p(\gamma \mid Z, I = i) = \frac{p(Z \mid \gamma)p(\gamma \mid I = i)}{p(Z \mid I = i)} \propto p(Z \mid \gamma)p(\gamma \mid I = i)$$

and

$$p(I = i \mid Z) = \frac{p_0(i) \cdot p(Z \mid I = i)}{\sum_{j=0}^{1} p_0(j) \cdot p(Z \mid j)}$$

We first consider the case $I = 0$, then we have

$$\begin{aligned}
p(Z \mid \gamma)p(\gamma \mid I = i) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\frac{(z - \gamma)^2}{\sigma^2}\right) \cdot \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{1}{2}\frac{(\gamma - \mu_1)^2}{\sigma_1^2}\right) \\
&= \frac{1}{2\pi\sigma_1\sigma} \exp\left(-\frac{1}{2}\left(\frac{z^2 + \gamma^2 - 2z\gamma}{\sigma^2} + \frac{\gamma^2 + \mu_1^2 - 2\gamma\mu_1}{\sigma_1^2}\right)\right) \\
&= \frac{1}{2\pi\sigma_1\sigma} \exp\left(-\frac{1}{2}\left(\frac{\sigma_1^2 z^2 + \sigma_1^2\gamma^2 - 2\sigma_1^2 z\gamma + \sigma^2\gamma^2 + \sigma^2\mu_1^2 - 2\sigma^2\gamma\mu_1}{\sigma^2\sigma_1^2}\right)\right) \\
&= \frac{1}{2\pi\sigma_1\sigma} \exp\left(-\frac{1}{2}\left(\frac{(\sigma_1^2 + \sigma^2)\gamma^2 + \sigma_1^2 z^2 + \sigma^2\mu_1^2 - 2(\sigma_1^2 z + \sigma^2\mu_1)\gamma}{\sigma^2\sigma_1^2}\right)\right)
\end{aligned}$$

Multiplying the numerator and the denominator with

$$\frac{1}{\sigma^4\sigma_1^2} + \frac{1}{\sigma^2\sigma_1^4} = \frac{\sigma^2 + \sigma_1^2}{\sigma^4\sigma_1^4}$$

we will obtain

$$\frac{1}{2\pi\sigma_1\sigma} \exp\left(-\frac{1}{2}\left(\frac{\frac{(\sigma_1^2+\sigma^2)^2}{\sigma^4\sigma_1^4}\gamma^2 + \frac{(\sigma_1^2 z^2+\sigma^2\mu_1^2)(\sigma_1^2+\sigma^2)}{\sigma^4\sigma_1^4} - \frac{2(\sigma_1^2 z+\sigma^2\mu_1)(\sigma_1^2+\sigma^2)}{\sigma^4\sigma_1^4}}{\frac{1}{\sigma^2} + \frac{1}{\sigma_1^2}}\right)\right)$$

We can take out the term $\frac{(\sigma_1^2+\sigma^2)^2}{\sigma^4\sigma_1^4}$ and the term $1/(\frac{1}{\sigma^2} + \frac{1}{\sigma_1^2}) = \frac{\sigma^2\sigma_1^2}{\sigma_1^2+\sigma^2}$ so the equation is equal to

$$\frac{1}{2\pi\sigma_1\sigma} \exp\left(-\frac{1}{2}\frac{\sigma^2\sigma_1^2}{\sigma_1^2 + \sigma^2} \cdot \frac{(\sigma_1^2 + \sigma^2)^2}{\sigma^4\sigma_1^4}\left(\gamma^2 + \frac{\sigma_1^2 z^2 + \sigma^2\mu_1^2}{\sigma_1^2 + \sigma^2} - 2\frac{(\sigma_1^2 z + \sigma^2\mu_1)}{\sigma_1^2 + \sigma^2}\gamma\right)\right) \tag{3.3}$$

Denote $\tilde{\mu}_1 = \frac{(\sigma_1^2 z+\sigma^2\mu_1)}{\sigma_1^2+\sigma^2} = \frac{z/\sigma^2+\mu_1/\sigma_1^2}{1/\sigma^2+1/\sigma_1^2}$ Notice that the equation 3.3 has the nice form of a quadratic term $(\gamma-\tilde{\mu}_1)^2$ after we add and subtract the term $\tilde{\mu}_1^2$. Also notice that in equation 3.3,

$$\frac{\sigma^2\sigma_1^2}{\sigma_1^2+\sigma^2}\cdot\frac{(\sigma_1^2+\sigma^2)^2}{\sigma^4\sigma_1^4} = \frac{\sigma_1^2+\sigma^2}{\sigma_1^2\sigma^2} = \frac{1}{\tilde{\sigma}_1^2}$$

where $\tilde{\sigma}_1^2 = 1/(\frac{1}{\sigma^2}+\frac{1}{\sigma_1^2})$ This is nice since we now have $\tilde{\sigma}_1^2$ and $\tilde{\mu}_1^2$. We can rearrange equation 3.3 and so the equation becomes

$$\frac{1}{\sqrt{2\pi}\tilde{\sigma}_1}\exp(-\frac{1}{2}\frac{1}{\tilde{\sigma}_1^2}(\gamma-\tilde{\mu}_1)^2)\cdot R \tag{3.4}$$

where R is the remaning term in the equation 3.3. We can write out the full expansion of R, which is

$$R = \frac{1}{\sqrt{2\pi}}\cdot\frac{\tilde{\sigma}_1}{\sigma\sigma_1}\exp\left(-\frac{1}{2}\frac{1}{\tilde{\sigma}_1^2}(\frac{\sigma_1^2 z^2+\sigma^2\mu_1^2}{\sigma_1^2+\sigma^2}-\tilde{\mu}_1^2)\right) \tag{3.5}$$

Notice that $\frac{\tilde{\sigma}_1}{\sigma\sigma_1} = \frac{\sigma\sigma_1}{\sqrt{\sigma^2+\sigma_1^2}}\cdot\sigma\sigma_1 = \frac{1}{\sqrt{\sigma^2+\sigma_1^2}}$

And the term within the exponential term in equation 3.5

$$\frac{1}{\tilde{\sigma}_1^2}(\frac{\sigma_1^2 z^2+\sigma^2\mu_1^2}{\sigma_1^2+\sigma^2}-\tilde{\mu}_1^2) \tag{3.6}$$

Substituting

$$\frac{1}{\tilde{\sigma}_1^2} = \frac{\sigma_1^2+\sigma^2}{\sigma_1^2\sigma^2}, \text{ and } \tilde{\mu}_1 = \frac{(\sigma_1^2 z+\sigma^2\mu_1)}{\sigma_1^2+\sigma^2}$$

in the equation 3.6, we discover that it is equal to

$$\begin{aligned}
&\frac{1}{\tilde{\sigma}_1^2}(\frac{\sigma_1^2 z^2+\sigma^2\mu_1^2}{\sigma_1^2+\sigma^2}-\tilde{\mu}_1^2)\\
&=\frac{\sigma_1^2+\sigma^2}{\sigma_1^2\sigma^2}\cdot\frac{\sigma_1^2 z^2+\sigma^2\mu_1^2}{\sigma_1^2+\sigma^2}-\frac{\sigma_1^2+\sigma^2}{\sigma_1^2\sigma^2}\cdot\frac{(\sigma_1^2 z+\sigma^2\mu_1)^2}{(\sigma_1^2+\sigma^2)^2}\\
&=\frac{z^2}{\sigma^2}+\frac{\mu_1^2}{\sigma_1^2}-\frac{(\sigma_1^2 z+\sigma^2\mu_1)^2}{\sigma_1^2\sigma^2(\sigma_1^2+\sigma^2)}\\
&=\frac{z^2\sigma_1^2(\sigma_1^2+\sigma^2)+\mu_1^2\sigma^2(\sigma_1^2+\sigma^2)-\sigma_1^4 z^2-\sigma^4\mu_1^2-2\sigma_1^2\sigma^2 z\mu}{\sigma^2\sigma_1^2}\\
&=\frac{(z-\mu_1)^2}{(\sigma^2+\sigma_1^2)}
\end{aligned}$$

Therefore, combing the two parts, we find that the remaining term has a nice form of

$$R = \frac{1}{\sqrt{2\pi}\tilde{\sigma}_1}\exp(-\frac{1}{2}\frac{(z-\mu_1)^2}{(\sigma^2+\sigma_1^2)}) = \varphi(z;\mu_1,\sigma^2+\sigma_1^2)$$

where $\varphi\left(z;\mu,\sigma^2\right)$ is the probability density function of the normal distribution $\mathrm{N}\left(\mu,\sigma^2\right)$: $\varphi\left(z;\mu,\sigma^2\right) = \left(\sqrt{2\pi\sigma^2}\right)^{-1}\exp\left\{-(z-\mu)^2/\left(2\sigma^2\right)\right\}$

It remains for us to solve for the term

$$p(I=i\mid Z) = \frac{p_0(i)\cdot p(Z\mid I=i)}{\sum_{j=0}^{1} p_0(j)\cdot p(Z\mid j)}$$

We notice that $p_0(i) = p_1$ when $I=0$ and

$$p(Z\mid I=i) = \int_\gamma p(Z,\gamma\mid I=i)d\gamma = \int_\gamma p(Z\mid I=i,\gamma)p(\gamma\mid I=i)d\gamma = \int_\gamma p(Z\mid\gamma)p(\gamma\mid I=i)d\gamma$$

after integrating over γ, we obtain $p(Z\mid I=i) = \varphi(z;\mu_1,\sigma^2+\sigma_1^2)$.

We can solve the case when $I=i$ by a similar approach and then combining them together with the rule $p(\gamma\mid Z) = \sum_{i=0}^{1} p(I=i\mid Z)\cdot p(\gamma\mid Z,I=i)$, we will obtain the desired form defined in the proposition. $\square$

In our setting, Z in Proposition 3.2 is just our GWAS data $(\hat{\gamma}, \hat{\Gamma})$. We now explain why the choice of mixture Gaussian can increase statistical efficiency. Proposition 3.2 makes it clear that when $\mu_1 = \mu_2 = 0$, $\mathbb{E}[\gamma \mid Z] = -\mathbb{E}[\gamma \mid -Z]$. This means that the point estimate in equation 2.5 remain the same even if you switch a pair of observations from $(\hat{\gamma}_j, \hat{\Gamma}_j)$ to $(-\hat{\gamma}_j, -\hat{\Gamma}_j)$. This is useful because the way alleles are coded in a GWAS is often arbitrary. Additionally, using a spike-and-slab implementation is crucial for efficiency. Suppse a single Gaussian prior is used, every SNP undergoes the same amount of multiplicative shrinkage, resulting in a constant scaling of the first estimating function in equation 2.5. On the other hand, with a spike-and-slab prior, each genetic instrument is selectively shrunk based on its strength, potentially leading to increased efficiency. This efficiency gain is likely more significant when the two components are more different (σ_1 and σ_2 are very different).

4. MULTI-DIMENSIONAL γ_j CASE

We now consider the case when $\gamma_j \in \mathbb{R}^k$ is a vector. Namely, we consider the case when there exists pleitropic effects where the jth snippet will affect our disease outcomes via k different risk factors. We first consider the case where we assert that the snippet affects the outcome only through these k risk factors. Our model then becomes

$$\Gamma_j = \gamma_j^T \beta,\ \Gamma_j \in \mathbb{R},\ \gamma_j \in \mathbb{R}^k,\ \beta \in \mathbb{R}^k$$

We assume that $\hat{\Gamma}_j \in \mathbb{R}$ and $\hat{\gamma}_j \in \mathbb{R}^k$ follows the following distribution

$$\begin{bmatrix} \hat{\Gamma}_j \\ \hat{\gamma}_j \end{bmatrix} \sim N\left(\begin{bmatrix} \Gamma_j \\ \gamma_j \end{bmatrix}, \Sigma_{YX}\Sigma\Sigma_{YX} \right) \tag{4.1}$$

where we denote

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \Sigma_{YX} = \mathrm{Diag}\left(\sigma_Y, \sigma_{X_1}, \sigma_{X_2}, \cdots, \sigma_{X_k}\right)$$

where $\Sigma_{11} \in \mathbb{R}$, $\Sigma_{22} \in \mathbb{R}^{k\times k}$ is invertible, and $\Sigma_{12} = \Sigma_{21}^T$

Remark 4.2. We assume that for different $(\hat{\Gamma}_j, \hat{\gamma}_j)$ and $(\hat{\Gamma}_i, \hat{\gamma}_i), i \neq j$ follow different covariance matrix. However, in this section, we drop the subscript j in the σ_Y, σ_X, Σ for notation ease.

Similar to the previous case, we write out the form of log-likelihood for each $\hat{\gamma}_j, \hat{\Gamma}_j$.

$$l_j(\beta, \gamma_j) = -\frac{1}{2}\left(\begin{bmatrix} \hat{\Gamma}_j - \gamma_j^T\beta \\ \hat{\gamma}_j - \gamma_j \end{bmatrix}^T \mathrm{Diag}(\frac{1}{\sigma_Y}, \frac{1}{\sigma_{X_1}}, \cdots, \frac{1}{\sigma_{X_k}})\Sigma^{-1}\,\mathrm{Diag}(\frac{1}{\sigma_Y}, \frac{1}{\sigma_{X_1}}, \cdots, \frac{1}{\sigma_{X_k}}) \begin{bmatrix} \hat{\Gamma}_j - \gamma_j^T\beta \\ \hat{\gamma}_j - \gamma_j \end{bmatrix} + \log(\det(\Sigma)) \right) \tag{4.3}$$

We denote

$$x = \begin{bmatrix} \hat{\Gamma}_j \\ \hat{\gamma} \end{bmatrix} - \begin{bmatrix} \gamma_j^T\beta \\ \gamma_j \end{bmatrix} = \begin{bmatrix} \hat{\Gamma}_j - \gamma_j^T\beta \\ \hat{\gamma}_j - \gamma_j \end{bmatrix} \text{ and } \boldsymbol{\Sigma} = \Sigma_{YX}\Sigma\Sigma_{YX}$$

Then by multivariate chain rule, the partial derivative in respect to γ_j is

$$\begin{aligned} \frac{\partial l(\beta,\gamma_j)}{\partial \gamma_j} &= \left(\frac{\partial}{\partial x} x^T \boldsymbol{\Sigma}^{-1} x \right) \frac{dx}{d\gamma_j} \\ &= -\frac{1}{2} 2x^T\boldsymbol{\Sigma}^{-1} \begin{bmatrix} -\beta^T \\ \mathbb{I} \end{bmatrix} \end{aligned}$$

We solve $\gamma_{j\,\mathrm{MLE}}$ by setting $\frac{\partial l(\beta,\gamma_j)}{\partial \gamma_j} = 0$ and so we have

$$\begin{aligned} &\frac{\partial l(\beta,\gamma_j)}{\partial \gamma_j} = 0 \\ &= x^T\boldsymbol{\Sigma}^{-1}\begin{bmatrix} \beta^T \\ \mathbb{I} \end{bmatrix} = 0 \\ &\to \left([\hat{\Gamma}_j, \hat{\gamma}_j^T] - [\gamma_j^T\beta, \gamma_j^T] \right) \boldsymbol{\Sigma}^{-1} \begin{bmatrix} \beta^T \\ \mathbb{I} \end{bmatrix} = 0 \end{aligned}$$

Rearranging the terms, we have

$$[\hat{\Gamma}_j, \hat{\gamma}_j^T]\mathbf{\Sigma}^{-1}\begin{bmatrix}\beta^T \\ \mathbb{I}\end{bmatrix} = [\gamma_j^T\beta, \gamma_j^T]\mathbf{\Sigma}^{-1}\begin{bmatrix}\beta^T \\ \mathbb{I}\end{bmatrix}$$

Taking transpose at both sides, we get

$$\begin{bmatrix}\beta & \mathbb{I}\end{bmatrix}\mathbf{\Sigma}^{-1}\begin{bmatrix}\hat{\Gamma}_j \\ \hat{\gamma}_j\end{bmatrix} = \begin{bmatrix}\beta & \mathbb{I}\end{bmatrix}\mathbf{\Sigma}^{-1}\begin{bmatrix}\beta^T \\ \mathbb{I}\end{bmatrix}\gamma_j$$

and therefore the MLE solution of γ_j can be written as

$$\gamma_{j\,\mathrm{MLE}} = \left(\begin{bmatrix}\beta, \mathbb{I}\end{bmatrix}\mathbf{\Sigma}^{-1}\begin{bmatrix}\beta^T \\ \mathbb{I}\end{bmatrix}\right)^{-1}\left(\begin{bmatrix}\beta, \mathbb{I}\end{bmatrix}\mathbf{\Sigma}^{-1}\begin{bmatrix}\hat{\Gamma}_j \\ \hat{\gamma}_j\end{bmatrix}\right) \tag{4.4}$$

Observe that this shares a very similar form of the equation 1.2 when $\gamma_j \in \mathbb{R}$ is one dimensional. Therefore, intuitively,

$$W_j(\beta) = \begin{bmatrix}\beta, \mathbb{I}\end{bmatrix}\mathbf{\Sigma}^{-1}\begin{bmatrix}\hat{\Gamma}_j \\ \hat{\gamma}_j\end{bmatrix}$$

is a sufficient statistics for γ_j

Similarly we can slove the partial derivative with respect to β. Specifically, we have

$$\begin{aligned}\frac{\partial l(\beta, \gamma_j)}{\partial \beta} &= (\frac{\partial}{\partial x}x^T\mathbf{\Sigma}^{-1}x)\frac{dx}{d\beta} \\ &= -\frac{1}{2}2x^T\mathbf{\Sigma}^{-1}\begin{bmatrix}-\gamma_j^T \\ 0\end{bmatrix} \\ &= (\begin{bmatrix}\hat{\Gamma}_j, \hat{\gamma}_j^T\end{bmatrix} - \begin{bmatrix}\gamma_j^T\beta, \gamma_j^T\end{bmatrix})\mathbf{\Sigma}^{-1}\begin{bmatrix}\gamma_j^T \\ 0\end{bmatrix}\end{aligned}$$

Definition 4.5. The score function $\psi(\beta)$ for $\beta \in \mathbb{R}^k$ is defined as the derivative of the profile log-likelihood over β. Namely,

$$\psi(\beta) = \frac{\partial l(\beta, \gamma_j)}{\partial \beta} = (\begin{bmatrix}\hat{\Gamma}_j, \hat{\gamma}_j^T\end{bmatrix} - \begin{bmatrix}\gamma_j^T\beta, \gamma_j^T\end{bmatrix})\mathbf{\Sigma}^{-1}\begin{bmatrix}\gamma_j^T \\ 0\end{bmatrix}$$

Then similarly to the one dimensional case, we define the conditional score function to be

Definition 4.6. The conditional score function is the residual of the score function $\psi(\beta)$ conditioning on $W_j(\beta)$, which is given by

$$\begin{aligned}C_j(\beta, \gamma_j) &= \psi(\beta, \gamma_j) - \mathbb{E}[\psi(\beta, \gamma_j) \mid W_j(\beta)] \\ &= \gamma_j^*\frac{(\hat{\Gamma}_j - \hat{\gamma}_j^T\beta)}{\mathrm{Var}(\hat{\Gamma}_j - \hat{\gamma}_j^T\beta)}\end{aligned}$$

where

$$\mathrm{Var}(\hat{\Gamma}_j - \hat{\gamma}_j^T\beta) = ([1, -\beta^T]\mathbf{\Sigma}\begin{bmatrix}1 \\ -\beta\end{bmatrix})^{-1} \text{ and } \gamma_j^* = \gamma_j^*(W_j(\beta))$$

Then, the total conditional score function is the aggregation of conditional score functions over all the jth snippets

Definition 4.7. The total conditional score function is defined as

$$C(\beta) = \sum_{j=1}^{p} C_j(\beta, W_j(\beta)) = \sum_{j=1}^{p}\gamma_j^*\frac{(\hat{\Gamma}_j - \hat{\gamma}_j^T\beta)}{\mathrm{Var}(\hat{\Gamma}_j - \hat{\gamma}_j^T\beta)}$$

To generalize the total conditional score function similar to definition 2.7, we can define

Definition 4.8. The generalized estimating equation for $\beta \in \mathbb{R}^k$ when $\gamma_j \in \mathbb{R}^k$ is defined as

$$C(\beta) = \sum_{j=1}^{p}\frac{f_j(W_j(\beta), \beta)\phi(t_j(\beta))}{\sqrt{Var(\hat{\Gamma}_j - \hat{\gamma}_j^T\beta)}}$$

where

$$t_j(\beta) = \frac{\hat{\Gamma}_j - \hat{\gamma}_j^T \beta}{\sqrt{Var(\hat{\Gamma}_j - \hat{\gamma}_j^T \beta)}}$$

Similarly, we can estimate γ_j^* through the empirical Bayes method, which will give us

$$\hat{\gamma}_{j\,\mathrm{EB}}^* = \mathbb{E}_{\pi(\hat{\eta})}[\gamma_j \mid W_j(\beta)]$$

We first consider a simple case when $\mathbf{\Sigma} = \Sigma_{YX}^2 = \mathrm{Diag}(\sigma_Y^2, \sigma_{X_1}^2, \cdots, \sigma_{X_k}^2)$

Similar to the one dimensional case, when $\hat{\Gamma}_j$ and $\hat{\gamma}_j$ are independent and $\hat{\gamma}_{jk}$ is independent to $\hat{\gamma}_{ji}$ for $k \neq i$, each entry of $W_j(\beta)$ is independent with the residual $\hat{\Gamma}_j - \hat{\gamma}_j^T \beta$

Lemma 4.9. *If $\mathbf{\Sigma} = \Sigma_{YX}^2 = \mathrm{Diag}(\sigma_Y^2, \sigma_{X_1}^2, \cdots, \sigma_{X_k}^2)$, each entry of $W_j(\beta)$ is independent with the residual $\hat{\Gamma}_j - \hat{\gamma}_j^T \beta$*

Proof. We first observe that if $\mathbf{\Sigma} = \mathrm{Diag}(\sigma_Y^2, \sigma_{X_1}^2, \cdots, \sigma_{X_k}^2)$, $W_j(\beta)$ asserts a nice form that is

$$\begin{aligned}
W_j(\beta) &= \begin{bmatrix}\beta, \mathbb{I}\end{bmatrix} \mathbf{\Sigma}^{-1} \begin{bmatrix}\hat{\Gamma}_j \\ \hat{\gamma}_j\end{bmatrix} \\
&= \begin{bmatrix} \beta_1 & 1 & 0 & \cdots & 0 \\ \beta_2 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \beta_k & 0 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} 1/\sigma_Y^2 & 0 & 0 & \cdots & 0 \\ 0 & 1/\sigma_{X_1}^2 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 1/\sigma_{X_k}^2 \end{bmatrix} \begin{bmatrix} \hat{\Gamma}_j \\ \hat{\gamma}_{j1} \\ \vdots \\ \hat{\gamma}_{jk} \end{bmatrix} \\
&= \begin{bmatrix} \frac{\hat{\Gamma}_j \beta_1}{\sigma_Y^2} + \frac{\hat{\gamma}_{j1}}{\sigma_{X1}^2} \\ \frac{\hat{\Gamma}_j \beta_2}{\sigma_Y^2} + \frac{\hat{\gamma}_{j2}}{\sigma_{X2}^2} \\ \vdots \\ \frac{\hat{\Gamma}_j \beta_k}{\sigma_Y^2} + \frac{\hat{\gamma}_{jk}}{\sigma_{Xk}^2} \end{bmatrix}
\end{aligned}$$

Then for each entry of $W_j(\beta)$, its covariance with the residual is equal to

$$\begin{aligned}
\mathrm{Cov}(\frac{\hat{\Gamma}_j \beta_k}{\sigma_Y^2} + \frac{\hat{\gamma}_{jk}}{\sigma_{Xk}^2}, \hat{\Gamma}_j - \hat{\gamma}_j^T \beta) &= \mathrm{Cov}(\frac{\hat{\Gamma}_j \beta_k}{\sigma_Y^2}, \hat{\Gamma}_j) - \mathrm{Cov}(\frac{\hat{\gamma}_{jk}}{\sigma_{Xk}^2}, \hat{\gamma}_j^T \beta) - \mathrm{Cov}(\frac{\hat{\Gamma}_j \beta_k}{\sigma_Y^2}, \hat{\gamma}_j^T \beta)) + \mathrm{Cov}(\frac{\hat{\gamma}_{jk}}{\sigma_{Xk}^2}, \hat{\Gamma}_j) \\
&= \frac{\beta_k}{\sigma_Y^2} \mathrm{Cov}(\hat{\Gamma}_j, \hat{\Gamma}_j) - \frac{\beta_k}{\sigma_{X_k}^2} \mathrm{Cov}(\hat{\gamma}_{jk}, \hat{\gamma}_{jk}) \\
&= 0
\end{aligned}$$

where the second equation holds because $\hat{\gamma}_j, \hat{\Gamma}_j$ are independent, and $\hat{\gamma}_{jk}$ is independent to $\hat{\gamma}_{ji}$ for $k \neq i$. Since they are multivariate Gaussian, covariance equal to 0 means they are independent □

With the previous lemma, we can then show that the expectation of the conditional score function $C_j(\beta)$ is equal to 0 at the true β_0.

Theorem 4.10. *The expectation of the conditional score function $C_j(\beta)$ is equal to 0 at the true β_0. Namely,*

$$\mathbb{E}[C_j(\beta_0)] = 0$$

Proof. We expand the formula for the conditional score function

$$\mathbb{E}[C_j(\beta_0)] = \mathbb{E}\left[\gamma_j^*(W_j(\beta_0)) \frac{(\hat{\Gamma}_j - \hat{\gamma}_j^T \beta_0)}{\mathrm{Var}(\hat{\Gamma}_j - \hat{\gamma}_j^T \beta_0)}\right]$$

By the previous lemma $W_j(\beta_0)$ and $\hat{\Gamma}_j - \hat{\gamma}_j^T \beta_0$ is independent, and so

$$\mathbb{E}[C_j(\beta_0)] = \mathbb{E}\left[\frac{\gamma_j^*(W_j(\beta_0))}{\mathrm{Var}(\hat{\Gamma}_j - \hat{\gamma}_j^T \beta_0)}\right] \mathbb{E}[(\hat{\Gamma}_j - \hat{\gamma}_j^T \beta_0)] = 0$$

where the last equation holds because $\mathbb{E}[\hat{\Gamma}_j] - \mathbb{E}[\hat{\gamma}_j^T]\beta_0 = \Gamma_j - \gamma_j^T \beta = 0$ □

Therefore, under the independence assumption specified in lemma 4.9, setting $C_j(\beta) = 0$ and solve for the optimal β is indeed a valid method.

Then, in order to obtain $\hat{\gamma}^*_{j\,\mathrm{EB}}$, we obtain $\hat{\gamma}^*_{jk\,\mathrm{EB}} = \mathbb{E}[\gamma_{jk} \mid W_{jk}(\beta)]$, where γ_{jk} and $W_{jk}(\beta)$ are the kth entry of γ_j and $W_j(\beta)$ respectively. This then recovers back to the one dimensional setting and we use proposition 3.2 to solve for the empirical bayes estimator for each γ_{jk}.

4.1. **What if $\Sigma \neq \mathbb{I}$.** We now focus on what would happen if $\Sigma \neq \mathbb{I}$, namely $\mathbf{\Sigma} = \Sigma_{YX}\Sigma\Sigma_{YX} \neq \Sigma^2_{YX}$. We want to show that even when $\mathbf{\Sigma} \neq \Sigma^2_{YX}$, $W_j(\beta)$ is independent with the residual $\hat{\Gamma}_j - \hat{\gamma}_j^T\beta$.

Lemma 4.11. *$W_j(\beta)$ is independent with the residual $\hat{\Gamma}_j - \hat{\gamma}_j^T\beta$.*

Proof. We first define $\mathbf{1} = [1, 1, \cdots, 1]^T \in \mathbb{R}^k$ to be the column vector in $\mathbb{R}^k$ where all entries are equal to 1. Then to prove that $W_j(\beta)$ is independent with the residual $\hat{\Gamma}_j - \hat{\gamma}_j^T\beta$, we can prove that $W_j(\beta) \perp\!\!\!\perp \mathbf{1}\cdot(\hat{\Gamma}_j - \hat{\gamma}_j^T\beta)$, and then prove that the covariance of $W_j(\beta)$ and $\mathbf{1}\cdot(\hat{\Gamma}_j - \hat{\gamma}_j^T\beta)$ is 0 since they are jointly Gaussian of the same dimension. Notice that

$$\hat{\Gamma}_j - \hat{\gamma}_j^T\beta = \begin{bmatrix} 1 & -\beta^T \end{bmatrix} \begin{bmatrix} \hat{\Gamma}_j \\ \hat{\gamma}_j \end{bmatrix}$$

We then write out the formula of the covariance to be

$$\begin{aligned}
\mathrm{Cov}(W_j(\beta), \mathbf{1}\cdot(\hat{\Gamma}_j - \hat{\gamma}_j^T\beta)) &= \mathrm{Cov}(\begin{bmatrix}\beta, \mathbb{I}\end{bmatrix} \mathbf{\Sigma}^{-1} \begin{bmatrix} \hat{\Gamma}_j \\ \hat{\gamma}_j \end{bmatrix}, \mathbf{1}\cdot \begin{bmatrix} 1 & -\beta^T \end{bmatrix} \begin{bmatrix} \hat{\Gamma}_j \\ \hat{\gamma}_j \end{bmatrix}) \\
&= \begin{bmatrix}\beta, \mathbb{I}\end{bmatrix} \mathbf{\Sigma}^{-1} \mathrm{Cov}(\begin{bmatrix} \hat{\Gamma}_j \\ \hat{\gamma}_j \end{bmatrix}, \begin{bmatrix} \hat{\Gamma}_j \\ \hat{\gamma}_j \end{bmatrix})(\mathbf{1}\cdot \begin{bmatrix} 1 & -\beta^T \end{bmatrix})^T \\
&= \begin{bmatrix}\beta, \mathbb{I}\end{bmatrix} \mathbf{\Sigma}^{-1}\mathbf{\Sigma} \begin{bmatrix} 1 \\ -\beta \end{bmatrix} \cdot \mathbf{1}^T \\
&= \begin{bmatrix}\beta, \mathbb{I}\end{bmatrix} \begin{bmatrix} 1 \\ -\beta \end{bmatrix} \cdot \mathbf{1}^T \\
&= 0
\end{aligned}$$

where the last equation holds because

$$\begin{bmatrix}\beta, \mathbb{I}\end{bmatrix} \begin{bmatrix} 1 \\ -\beta \end{bmatrix} = \beta - \mathbb{I}\cdot\beta = 0$$

□

Therefore, from Lemma 4.11, we can see that even if when $\mathbf{\Sigma}$ is an arbitrary covariance matrix, $W_j(\beta)$ is still independent with the residual $\hat{\Gamma}_j - \hat{\gamma}_j^T\beta$. This means that Theorem 4.10 still holds when $\hat{\Gamma}_j$ and $\hat{\gamma}_j$ have arbitrary dependence structure. Therefore, the conditional score function is still a valid method to solve for the optimal β^*.

4.2. **pleitropic case.** We now consider the case when there exists more pleitropic effect when $1, 2, \cdots, k$ risk factors cannot cover all the effects of the jth snippet on the disease outcome. Adding upon equation 4.12, Our model then becomes

$$\begin{bmatrix} \hat{\Gamma}_j \\ \hat{\gamma}_j \end{bmatrix} \sim N\left(\begin{bmatrix} \Gamma_j \\ \gamma_j \end{bmatrix}, \Sigma_{YX}\Sigma\Sigma_{YX} + \tau\mathbf{1}\mathbf{1}^T \right) \tag{4.12}$$

where $\tau \in \mathbb{R}$ is some unknown noise, and $\mathbf{1} = [1, 1, \cdots, 1] \in \mathbb{R}^{p+1}$

To simplify the problem, we will obtain $\hat{\gamma}^*_{jk}$ by obtaining the empirical bayes estimator by conditioning on each $W_{jk}(\beta, \tau^2)$ which is the kth entry of (β, τ^2). We can then apply Proposition 3.2 to obtain the estimator $\hat{\gamma}^*_{jk} = \mathbb{E}[\hat{\gamma}^*_{jk} \mid W_{jk}(\beta, \tau^2)]$.

References

[1] Qingyuan Zhao, Jingshu Wang, Gibran Hemani, Jack Bowden, and Dylan Small. Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score. https://projecteuclid.org/journals/annals-of-statistics/volume-48/issue-3/Statistical-inference-in-two-sample-summary-data-Mendelian-randomization-using/10.1214/19-AOS1866.full

[2] Qingyuan Zhao, Yang Chen, Jingshu Wang, Dylan S Small. Powerful three-sample genome-wide design and robust statistical inference in summary-data Mendelian randomization, International Journal of Epidemiology, Volume 48, Issue 5, October 2019, Pages 1478–1492, https://doi.org/10.1093/ije/dyz142