Variance Components Analysis
Course project

Zheng Cong

May 2020

1 Problem

We need to use linear mixed model to estimate the heritability of each of the four phenotypes.

$$\mathrm{y} = \mathrm{X}\beta + \mathrm{Wu} + \mathrm{e}$$

where y is a the $n \times 4$ genotype matrix.(Sample size $n = 5123$,$p = 319147$ is the number of genetic marker). $\mathbf{W} \in \mathbf{R}^{n\times p}$ is the standardized genotype matrix with zero mean and unit variance, and $\mathbf{e} \sim \mathcal{N}\left(0, \sigma_\epsilon^2 \mathbf{I}\right)$. $\mathbf{u} \sim \mathcal{N}\left(0, \frac{\sigma_u^2}{p}\mathbf{I}\right) \in \mathbf{R}^{p\times 1}$ is the coefficient corresponding to the fixed effect. And $\mathbf{X} \in \mathbf{R}^{n\times(10+1)}$ includes the principal component scores corresponding to the first ten leading principal components and a column of ones, $\beta \in \mathbf{R}^{11}$ is the coefficient of fixed effect.(Code available at ”” including all the experiment mentioned in the report.)

2 Estimate the model's parameter

2.1 EM algorithm

We could derive an EM algorithm to obtain an estimation of model parameters $\theta = \left\{\beta, \sigma_u^2, \sigma_\epsilon^2\right\}$.

$$\mathrm{y} = \mathrm{X}\beta + \mathrm{Wu} + \mathrm{e} \tag{1}$$

The complete-data log-likelihood is given as

$$\begin{aligned} \mathcal{L} &= \log \operatorname{Pr}(\mathbf{y}, u|\theta) \\ &= -\frac{n}{2}\log\left(2\pi\sigma_\epsilon^2\right) - \frac{1}{2\sigma_\epsilon^2}\|\mathbf{y} - \mathbf{X}\beta - \mathbf{W}u\|^2 \\ &\quad -\frac{p}{2}\log\left(2\pi\sigma_u^2\right) - \frac{p}{2\sigma_\beta^2}\|u\|^2 \end{aligned} \tag{2}$$

And the posterior of u is a Gaussian, denote it as $\mathcal{N}(u|m, \Sigma)$, where

$$\Sigma^{-1} = \frac{1}{\sigma_\epsilon^2}\mathbf{W}^T\mathbf{W} + \frac{p}{\sigma_u^2}\mathbf{I}_p$$

$$m=\frac{1}{\sigma_\epsilon^2}\Sigma\mathbf{W}^T(y-\mathbf{X}\beta)$$

Now, in the E-step, taking expectation w.r.t the posterior $\mathcal{N}(u|m,\Sigma)$. Denote $\hat{y}=y-X\beta$, then

$$\mathbf{E}\left[\|\hat{y}-\mathbf{W}u\|^2\right]=\hat{y}^T\hat{y}-2\hat{y}^TWm+m^TW^TWm+trW^TW\Sigma$$

$$\mathbf{E}\left[\|u\|^2\right]=m^Tm+\operatorname{tr}(\mathbf{\Sigma}))$$

Then the Q -function given the current estimates θ_{old} is obtained as:

$$\begin{aligned}\mathcal{Q}\left(\boldsymbol{\theta}|\boldsymbol{\theta}_{old}\right)=&-\frac{n}{2}\log\left(2\pi\sigma_\epsilon^2\right)-\frac{p}{2}\log\left(\frac{2\pi\sigma_u^2}{p}\right)\\&-\frac{1}{2\sigma_e^2}\|\mathbf{y}-\mathbf{X}\beta-\mathbf{W}\boldsymbol{m}\|^2-\frac{p}{2\sigma_u^2}\boldsymbol{m}^T\boldsymbol{m}\\&-\operatorname{tr}\left(\left(\frac{1}{2\sigma_e^2}\mathbf{W}^T\mathbf{W}+\frac{p}{2\sigma_u^2}\mathbf{I}_p\right)\Sigma\right)\end{aligned}\tag{3}$$

In the M-step, the new estimates of the parameter θ is obtained by setting the corresponding derivative of the Q-function to be zero. The updating function is:

$$\sigma_e^2=\frac{1}{n}\left[\|\mathbf{y}-\mathbf{X}\beta-\mathbf{W}\boldsymbol{m}\|^2+\operatorname{tr}\left(\mathbf{W}^T\mathbf{W}\mathbf{\Sigma}\right)\right]$$

$$\sigma_\beta^2=\boldsymbol{m}^T\boldsymbol{m}+\operatorname{tr}(\mathbf{\Sigma})$$

$$\beta=(X^TX)^{-1}X^T(y-Wm)$$

In order to check convergence of the EM algorithm, We also need to evaluate the

$$ELBO=Q(\theta^{old})+0.5\times log|2\pi\Sigma|$$

2.2 PX-EM algorithm

Up to now, we have derived a whole EM algorithm. Now I will introduce another algorithm PX-EM[1], which is an extension of EM algorithm with a faster speed. And in the experiment, I use PX-EM to get an estimation of the variance component.

$$\mathrm{y}=\mathrm{X}\beta+\delta\mathrm{Wu}+\mathrm{e}\tag{4}$$

For the PX-EM, E-step is the same with EM algorithm. For the M-step, in addition to updating the parameters mentioned above, we also need to update parameter δ to speed up the algorithm. The resulting updates are given as follows,

$$\delta=\frac{(\mathrm{y}-\mathrm{X}\beta)^T\mathrm{W}m}{m^T\mathrm{W}^T\mathrm{W}m+\operatorname{tr}\left(\mathrm{W}^T\mathrm{W}\Sigma\right)}$$

$$\sigma_e^2=\frac{1}{n}\left[\|\mathbf{y}-\mathbf{X}\beta-\delta\mathbf{W}\boldsymbol{m}\|^2+\delta^2\operatorname{tr}\left(\mathbf{W}^T\mathbf{W}\mathbf{\Sigma}\right)\right]$$

$$\sigma_\beta^2 = \boldsymbol{m}^T\boldsymbol{m} + \mathrm{tr}(\boldsymbol{\Sigma})$$

$$\beta = (X^T X)^{-1} X^T (y - \delta W m)$$

The reduction step is to rescale $sigma_u^2 = \delta^2 sigma_u^2$ and reset $\delta = 1$

2.3 Matrix inverse lemma

As we can see, we have to obtain an inverse of an $p \times p$ matrix in the E-step. However, when p is very large, this process is not practical. So we have to use matrix inverse lemma

$$(W^T W + I_p)^{-1} W^T = W^T (W W^T + I_n)^{-1}$$

So the inverse of a $p \times p$ matrix is converted to a $n \times n$ inversion process. When $p >> n$, it will speed up the algorithm. Now, we derive PX-EM using this Lemma. First, we denote $WW^T = UDU^T$, where $D = diag(d_1, \ldots, d_n)$

E-step of the PX-EM becomes:

$$\Sigma^{-1} = \frac{1}{\sigma_\epsilon^2}\mathbf{W}^T\mathbf{W} + \frac{p}{\sigma_u^2}\mathbf{I}_p = \hat{(}D)$$

Denote $\hat{d} = diag(\hat{D}) = [\hat{d_1}, \ldots, \hat{d_n}, \hat{d_{n+1}}, \ldots, \hat{d_p}]$ and $\hat{d}1 = [\hat{d_1}, \ldots, \hat{d_n}]$, where $\hat{d_i} = \frac{p}{\sigma_u^2} + \frac{d_i}{\sigma_\epsilon^2}$ when $i <= n$ and $\hat{d_i} = \frac{p}{\sigma_u^2}$ when $i > n$.

$$m = \frac{1}{\sigma_\epsilon^2} W^T U[\mathbf{U}^T(y - \mathbf{X}\beta) \odot 1/\hat{d}1]$$

M-step:update the model parameters by

$$\delta = \frac{(\mathrm{y} - \mathrm{X}\beta)^T \mathrm{W} m}{m^T \mathrm{W}^T \mathrm{W} m + \sum_{i=1}^n \frac{d_i}{\hat{d_i}}}$$

$$\sigma_e^2 = \frac{1}{n}\left[\|\mathbf{y} - \mathbf{X}\beta - \delta\mathbf{W}\boldsymbol{m}\|^2 + \delta^2 \sum_{i=1}^n \frac{d_i}{\hat{d_i}}\right]$$

$$\sigma_\beta^2 = \boldsymbol{m}^T\boldsymbol{m} + \sum_{i=1}^p \frac{1}{\hat{d_i}}$$

$$\beta = (X^T X)^{-1} X^T (y - \delta W m)$$

2.4 Method of Moments

We could derive a MoM[2] estimator to check whether the result obtained by PX-EM is reasonable or not. The code about MoM is also included in my code. And the specific algorithm will be introduced in the section 4.

2.5 Fisher Information Matrix and Delta Method

The covariance of the variance component $\sigma_u^2, \sigma_\epsilon^2$ can be obtained from inverse of the Fisher Information Matrix.

The incomplete-data likelihood is

$$p(y|\theta) = \mathbf{N}(y|X\beta, \sigma_\epsilon^2 I_n + WW^n \frac{\sigma_u^2}{p})$$

, denote $\Omega = \sigma_\epsilon^2 I_n + WW^T \frac{\sigma_u^2}{p}$ and $K = WW^T/p$

The first derivative is

$$\frac{\partial \mathcal{L}}{\partial \sigma_u^2} = \frac{1}{2} \operatorname{tr} \left[-\Omega^{-1}\mathbf{K} + (\mathbf{y} - \mathbf{X}\beta)^T \Omega^{-1} \mathbf{K} \mathbf{\Omega}^{-1} (\mathbf{y} - \mathbf{X}\beta) \right]$$

$$\frac{\partial \mathcal{L}}{\partial \sigma_\epsilon^2} = \frac{1}{2} \operatorname{tr} \left[-\Omega^{-1} + (\mathbf{y} - \mathbf{X}\beta)^T \Omega^{-2} (\mathbf{y} - \mathbf{X}\beta) \right]$$

The second derivative is

$$\frac{\partial^2 \mathcal{L}}{\partial \left(\sigma_u^2\right)^2} = \frac{1}{2} \operatorname{tr} \left[\left(\Omega^{-1}\mathbf{K}\right)^2 - 2\left(\Omega^{-1}\mathbf{K}\right)^2 \Omega^{-1} (\mathbf{y} - \mathbf{X}\beta)(\mathbf{y} - \mathbf{X}\beta)^T \right]$$

$$\frac{\partial^2 \mathcal{L}}{\partial \left(\sigma_\epsilon^2\right)^2} = \frac{1}{2} \operatorname{tr} \left[\Omega^{-2} - 2\Omega^{-3} (\mathrm{y} - \mathrm{X}\beta)(\mathrm{y} - \mathrm{X}\beta)^T \right]$$

$$\frac{\partial^2 \mathcal{L}}{\partial \sigma_u^2 \partial \sigma_\epsilon^2} = \frac{1}{2} \operatorname{tr} \left[\Omega^{-1}\mathbf{K}\Omega^{-1} - \left(\Omega^{-1}\mathbf{K}\Omega^{-2} + \Omega^{-2}\mathbf{K}\Omega^{-1}\right)(\mathbf{y} - \mathbf{X}\beta)(\mathbf{y} - \mathbf{X}\beta)^T \right]$$

And $E\left[(\mathbf{y} - \mathbf{X}\beta)(\mathbf{y} - \mathbf{X}\beta)^T\right] = \Omega$,so the Fisher Information Matrix can be represented as

$$FIM = \frac{1}{2} \begin{bmatrix} \operatorname{tr}\left[\left(\Omega^{-1}\mathbf{K}\right)^2\right] & \operatorname{tr}\left(\Omega^{-1}\mathbf{K}\Omega^{-1}\right) \\ . & \operatorname{tr}\left[\Omega^{-2}\right] \end{bmatrix}$$

And the covariance matrix of the variance component is the inverse of FIM.

Denote the covariance of the variance component is $\operatorname{cov}(\sigma_u^2, \sigma_\epsilon) = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}$ Then the standard error of heritability is

$$se = \sigma_{11}\left(\frac{\sigma_\epsilon^2}{(\sigma_u^2 + \sigma_\epsilon^2)^2}\right)^2 - 2\sigma_{12}\frac{\sigma_\epsilon^2 \sigma_u^2}{(\sigma_\epsilon^2 + \sigma_u^2)^4} + \sigma_{22}\left(\frac{\sigma_u^2}{(\sigma_u^2 + \sigma_\epsilon^2)^2}\right)^2$$

2.6 Experiment Result

The estimation of model parameters for the first phenotype is $\theta = \left\{\sigma_u^2, \sigma_\epsilon^2\right\} = [0.2264, 0.7640]$ The heritability of the first phenotype is 0.226. Covariance Matrix is $cov(\sigma_u^2, \sigma_\epsilon^2) = [[0.00266937 - 0.00248562], [-0.002485620.00268829]]$ Standard error of the heritability is 0.00266

The estimation of model parameters for the second phenotype is $\theta = \left\{\sigma_u^2, \sigma_\epsilon^2\right\} = [0.3040, 0.6898]$ The heritability of the second phenotype is 0.304. Covariance Matrix is $cov(\sigma_u^2, \sigma_\epsilon^2) = [[0.00280057 - 0.00255191], [-0.002551910.002695]]$ Standard error of the heritability is 0.00272

The estimation of model parameters for the third phenotype is $\theta = \left\{\sigma_u^2, \sigma_\epsilon^2\right\} = [0.2929, 0.6916]$ The heritability of the third phenotype is 0.293. Covariance Matrix is $cov(\sigma_u^2, \sigma_\epsilon^2) = [[0.00273743 - 0.00250017], [-0.002500170.00264705]]$ Standard error of the heritability is 0.00272

The estimation of model parameters for the fourth phenotype is $\theta = \left\{\sigma_u^2, \sigma_\epsilon^2\right\} = [0.1655, 0.8325]$ The heritability of the fourth phenotype is 0.165. Covariance Matrix is $cov(\sigma_u^2, \sigma_\epsilon^2) = [[0.00259856 - 0.00246401], [-0.002464010.00272019]]$ Standard error of the heritability is 0.00257

3 Other Method to estimate σ_ϵ^2

3.1 Scaled Lasso

Scaled Lasso is a method which jointly estimates the regression coefficients and noise level in a linear model. More specifically, we need to minimize the joint loss function

$$L_{\lambda_0}(\beta, \sigma_\epsilon) = \frac{\|y - X\beta\|_2^2}{2n\sigma_\epsilon} + \frac{(1-a)\sigma_\epsilon}{2} + \lambda_0|\beta|_1$$

, where λ_0 is a constant, σ_ϵ represents the noise level. The algorithm is just like this,

Scaled-Lasso algorithm
Initialization: $\widehat{\beta} = 0$
Repeat until convergence
- update $\widehat{\sigma} : \widehat{\sigma} = \|Y - \mathbf{X}\widehat{\beta}\|_2^2/\sqrt{n}$
- update the model's coefficients $\widehat{\beta} : \min_\beta \frac{\|y - X\beta\|^2}{n} + \widehat{\sigma}\lambda_0|\beta|_1$
Output: $\widehat{\beta}$

For the scaled lasso algorithm, the estimation of σ_ϵ is almost equal to the variance of the corresponding phenotype.

4 Estimate The Correlation Between Different Phenotypes

4.1 Method of moments

For this part, we use MoM to get an estimation of model parameter $\theta = \sigma_{u_1}^2, \sigma_{e_1}^2, \sigma_{u_2}^2, \sigma_{e_2}^2, \rho, \rho_e$

Consider the following linear mixed model to jointly model two phenotypes

$$\mathbf{y}_1 = \mathbf{X}\boldsymbol{\beta}_1 + \mathbf{W}\mathbf{u}_1 + \mathbf{e}_1, \mathbf{y}_2 = \mathbf{X}\boldsymbol{\beta}_2 + \mathbf{W}\mathbf{u}_2 + \mathbf{e}_2 \tag{5}$$

Denote $\hat{\rho} = \rho\sigma_{u_1}\sigma_{u_2}$, $\hat{\rho}_e = \rho_e\sigma_{u_1}\sigma_{u_2}$, $y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$, $K = WW^T$,
$\boldsymbol{\Gamma}_u = \begin{bmatrix} \sigma_{u_1}^2 & \hat{\rho} \\ \hat{\rho} & \sigma_{u_2}^2 \end{bmatrix}$, $\boldsymbol{\Gamma}_e = \begin{bmatrix} \sigma_{e_1}^2 & \hat{\rho}_e \\ \hat{\rho}_e & \sigma_{e_2}^2 \end{bmatrix}$
then

$$\begin{pmatrix} u_{1,j} \\ u_{2,j} \end{pmatrix} \sim \mathcal{N}\begin{pmatrix} \sigma_{u_1}^2 & \hat{\rho} \\ \hat{\rho} & \sigma_{u_2}^2 \end{pmatrix}, j = 1, \ldots, p, \quad \begin{pmatrix} e_{1,i} \\ e_{2,i} \end{pmatrix} \sim \mathcal{N}\begin{pmatrix} \sigma_{e_1}^2 & \hat{\rho}_e \\ \hat{\rho}_e & \sigma_{e_2}^2 \end{pmatrix}, i = 1, \ldots, n$$

First, we multiply equation (5) by the projection matrix $M = I_n - X(X^TX)^{-1}X^T$ to match the first moment. Then we only need to match the second moment, which means we need to solve the following ordinary least squares problem:

$$\text{argmin}_\theta \|(Ny)(Ny)^T - (\Gamma_u \otimes MKM + \Gamma_e \otimes M)\|_F^2$$

where $N = \begin{bmatrix} M & 0 \\ 0 & M \end{bmatrix}$

It can be rewritten as

$$\text{tr}[((Ny)(Ny)^T - (\Gamma_u \otimes MKM + \Gamma_e \otimes M))^2]$$

=(Ny)(Ny)$^T(Ny)(Ny)^T + (\sigma_{u_1}^4 + 2\hat{\rho}^2 + \sigma_{u_2}^4)\,\text{tr}(MKMK) + (\sigma_{e_1}^4 + 2\hat{\rho_e}^2 + \sigma_{e_2}^4)\,\text{tr}(M)$

-2$(\sigma_{u_1}^2 y_1^T MKMy_1 + 2\hat{\rho} y_2^T MKMy_1 + \sigma_{u_2}^2 y_2^T MKMy_2) - 2(\sigma_{e_1}^2 y_1^T My_1 + 2\hat{\rho}_e y_2^T My_1 + \sigma_{e_2}^2 y_2^T My_2) + 2(\sigma_{u_1}^2\sigma_{e_1}^2 + 2\hat{\rho}\hat{\rho}_e + \sigma_{u_2^2}\sigma_{e_2^2})\,\text{tr}(MK)$

which leads to a normal equation

$$S\theta = q$$

$$\begin{bmatrix} \text{tr}\left[(MK)^2\right] & \text{tr}\left[MK\right] & 0 & 0 & 0 & 0 \\ \text{tr}\left[MK\right] & \text{tr}\left[M\right] & 0 & 0 & 0 & 0 \\ 0 & 0 & \text{tr}\left[(MK)^2\right] & \text{tr}\left[MK\right] & 0 & 0 \\ 0 & 0 & \text{tr}\left[MK\right] & \text{tr}\left[M\right] & 0 & 0 \\ 0 & 0 & 0 & 0 & \text{tr}\left[(MK)^2\right] & \text{tr}\left[MK\right] \\ 0 & 0 & 0 & 0 & \text{tr}\left[MK\right] & \text{tr}\left[M\right] \end{bmatrix} \begin{bmatrix} \sigma_{u_1}^2 \\ \sigma_{e_1}^2 \\ \sigma_{u_2}^2 \\ \sigma_{e_2}^2 \\ \hat{\rho} \\ \hat{\rho}_e \end{bmatrix} = \begin{bmatrix} \mathrm{y}_1^T MKM\mathrm{y}_1 \\ \mathrm{y}_1^T M\mathrm{y}_1 \\ \mathrm{y}_2^T MKM\mathrm{y}_2 \\ \mathrm{y}_2^T M\mathrm{y}_2 \\ \mathrm{y}_2^T MKM\mathrm{y}_1 \\ \mathrm{y}_2^T M\mathrm{y}_1 \end{bmatrix}$$

We can get an estimation of β by solving the normal equation.

4.2 Experiment and Results

Denote $\rho(i,j)$ is the correlation between i'th phenotype and j'th phenotype, then the result is

$$\rho(0,1) = 0.264, \rho_e(0,1) = 0.070$$

$$\rho(0,2) = 0.948, \rho_e(0,2) = 0.855$$

$$\rho(0,3) = 0.301, \rho_e(0,3) = 0.455$$

$$\rho(1,2) = -0.057, \rho_e(1,2) = -0.176$$

$$\rho(1,3) = -0.398, \rho_e(1,3) = -0.410$$

$$\rho(2,3) = 0.313, \rho_e(2,3) = 0.346$$

From experiment result, we can see 0'th phenotype and 2'th phenotype are highly correlated.

References

[1] Chuanhai Liu, Donald B Rubin, and Ying Nian Wu. Parameter expansion to accelerate em: the px-em algorithm. *Biometrika*, 85(4):755–770, 1998.

[2] Yue Wu and Sriram Sankararaman. A scalable estimator of snp heritability for biobank-scale data. *Bioinformatics*, 34(13):i187–i194, 2018.