

The computation of four AREs

Yuan Xue

xueyuan115@mails.ucas.ac.cn

August 11, 2018

1 Introduction

1.1 Notation

likelihood function

$$L_n(\lambda_1, \lambda_2, \eta) = \frac{\lambda_1^{r_1} \lambda_2^{r_2} \exp(r\eta)}{\{1 + \exp(\eta)\}^{n_0} \{1 + \lambda_1 \exp(\eta)\}^{n_1} \{1 + \lambda_2 \exp(\eta)\}^{n_2}} \quad (1)$$

log-likelihood function $l_n(\lambda, \eta, \theta)$

we work with $l_n(\lambda, 1 - \theta - \theta\lambda, \eta)$, where θ is explicitly expressed.

$$\begin{aligned} x_1 &= \lambda \\ x_2 &= 1 - \theta - \theta\lambda \\ x_3 &= \eta \end{aligned}$$

Denote

$$\begin{aligned} l_{n,\mu} &= \partial l_n / \partial x_\mu \quad \text{for } \mu = 1, 2, 3 \\ l_{n,\mu\nu} &= \partial^2 l_n / \partial x_\mu \partial x_\nu \quad \text{for } \mu = 1, 2, \nu = 1, 2, 3 \\ l_{n,33} &= \partial^2 l_n / \partial x_3 \partial x_3 \\ l_{n,\mu\nu} &= l_{n,\nu\mu} \quad \text{for } \mu, \nu = 1, 2 \\ l_{n,\mu\nu} &= l_{n,\nu\mu}^T \quad \text{for } \mu = 1, 2, \nu = 3 \\ L_{\mu\nu}(\eta) &= E_{H_0}(l_{1,\mu\nu}(1, 1, \eta)) \quad \text{for } \mu = 1, 2, 3, \nu = 1, 2, 3 \end{aligned}$$

$$s(\theta, \eta) = l_{1,1}(1, 1, \eta) + \theta l_{1,2}(1, 1, \eta) - (L_{13}^T(\eta) + \theta L_{23}^T(\eta)) L_{33}^{-1}(\eta) l_{1,3}(1, 1, \eta)$$

log-likelihood function

$$l(x_1, x_2, x_3) = r_1 \ln(x_1) + r_2 \ln(x_2) + r x_3 - n_0 \ln(1 + e^{x_3}) - n_1 \ln(1 + x_1 e^{x_3}) - n_2 \ln(1 + x_2 e^{x_3})$$

$$\begin{aligned} l_{n,1} &= \frac{r_1}{x_1} - \frac{n_1 e^{x_3}}{1 + x_1 e^{x_3}} \\ l_{n,2} &= \frac{r_2}{x_2} - \frac{n_2 e^{x_3}}{1 + x_2 e^{x_3}} \\ l_{n,3} &= r - \frac{n_0 e^{x_3}}{1 + e^{x_3}} - \frac{n_1 x_1 e^{x_3}}{1 + x_1 e^{x_3}} - \frac{n_2 x_2 e^{x_3}}{1 + x_2 e^{x_3}} \end{aligned}$$

$$\begin{aligned}
l_{n,11} &= -\frac{r_1}{x_1^2} + \frac{n_1 e^{2x_3}}{(1+x_1 e^{x_3})^2} \\
l_{n,12} &= 0 \\
l_{n,13} &= -\frac{n_1 e^{x_3}}{(1+x_1 e^{x_3})^2} \\
l_{n,22} &= -\frac{r_2}{x_2^2} + \frac{n_2 e^{2x_3}}{(1+x_2 e^{x_3})^2} \\
l_{n,23} &= -\frac{n_2 e^{x_3}}{(1+x_2 e^{x_3})^2} \\
l_{n,33} &= -\frac{n_0 e^{x_3}}{(1+e^{x_3})^2} - \frac{n_1 x_1 e^{x_3}}{(1+x_1 e^{x_3})^2} - \frac{n_2 x_2 e^{x_3}}{(1+x_2 e^{x_3})^2}
\end{aligned}$$

1.2 Algorithm

- Input: $Y, G, \theta^{(0)}, \theta_i, \theta_j$
- Output: $e_P(Z_{MERT}, Z_{\theta^{(0)}}), \tilde{e}_C(Z_{MERT}, Z_{\theta^{(0)}}), e_{HL}(Z_{MERT}, Z_{\theta^{(0)}}), e_B(Z_{MERT}, Z_{\theta^{(0)}})$

step 1 Estimate $\hat{\eta}$. where $\hat{\eta}$ satisfy $\partial l_n / \partial \eta|_{H_0, \hat{\eta}_n} = l_{n,3}(1, 1, \hat{\eta}_n) = 0$.

step 2 Compute $l_n(1, 1, \hat{\eta}); l_{n,\mu}(1, 1, \hat{\eta})$, for $\mu = 1, 2, 3$; $l_{n,\mu\nu}(1, 1, \hat{\eta})$ for $\mu = 1, 2, 3, \nu = 1, 2, 3$.

step 3 Compute $L_{\mu\nu}(\hat{\eta}) = E_{H_0}(l_{1,\mu\nu}(1, 1, \hat{\eta})) = \frac{1}{n} l_{n,\mu\nu}(1, 1, \hat{\eta})$

step 4 Compute $\sigma(\theta^{(0)}), \sigma(\theta_i), \sigma(\theta_j), \sigma(\theta^{(0)}, \theta_i), \sigma(\theta^{(0)}, \theta_j), \sigma(\theta_i, \theta_j)$, where

$$\sigma(\theta_i, \theta_j) = A_{\hat{\eta}} \theta_i \theta_j + B_{\hat{\eta}}(\theta_i + \theta_j) + C_{\hat{\eta}}$$

$$A_{\eta} = L_{23}(\eta) L_{33}^{-1}(\eta) L_{32}(\eta) - L_{22}(\eta)$$

$$B_{\eta} = L_{13}(\eta) L_{33}^{-1}(\eta) L_{31}(\eta) - L_{12}(\eta)$$

$$C_{\eta} = L_{13}(\eta) L_{33}^{-1}(\eta) L_{31}(\eta) - L_{11}(\eta)$$

step 5 1). Given (λ, θ) , for $i = 1, \dots, n$ (typically $n \geq 10,000$), sample x_i from a given distribution (see ZLY for detail), then given this x_i , sample y_i from $f(y|\lambda, 1 - \theta + \theta\lambda, \eta_0^T x_i)$ (see ZLY for how to choose the density $f(\cdot|\cdot, \cdot, \cdot)$). Then we get a sample $\{(y_i, x_i) : i = 1, \dots, n\}$ under H_1 .

2). Set

$$\mu(\lambda, \theta) = E_{H_1, \eta_0}(s(\theta, \eta_{\theta}))$$

$$\approx \frac{1}{n} \sum_{i=1}^n \left(l_{1,11}(1, 1, \hat{\eta}|y_i, x_i) + \theta l_{1,21}(1, 1, \hat{\eta}|y_i, x_i) - (L_{13}^T(\hat{\eta}) + \theta L_{23}^T(\hat{\eta})) L_{33}^{-1}(\hat{\eta}) l_{1,13}(1, 1, \hat{\eta}|y_i, x_i) \right).$$

step 6 Compute $\mu(\lambda, \theta^{(0)}), \mu(\lambda, \theta_i), \mu(\lambda, \theta_j)$,

For fixed (λ, θ) , from the P.5, η_{θ} is consistently estimated by $\hat{\eta}_n$, so we simulate $n^{H_1} = 2000$ samples with fixed (λ, θ) under H_1 , then calculate the $\mu(\lambda, \theta)$ and its derivatives.

$$\mu(\lambda, \theta) = E_{H_1, \eta_0}(s(\theta, \eta_{\theta})) = \frac{1}{n^{H_1}} (l_{n,1}^{H_1}(1, 1, \hat{\eta}) + \theta l_{n,2}^{H_1}(1, 1, \hat{\eta}) - (L_{13}^T(\hat{\eta}) + \theta L_{23}^T(\hat{\eta})) L_{33}^{-1}(\hat{\eta}) l_{n,3}^{H_1}(1, 1, \hat{\eta}))$$

step 7 Compute $\mu^{(1)}(\lambda, \theta^{(0)}), \mu^{(1)}(\lambda, \theta_i), \mu^{(1)}(\lambda, \theta_j)$,

$$\mu^{(1)}(\lambda, \theta) = \frac{1}{n^{H_1}} (l_{n,11}^{H_1}(1, 1, \hat{\eta}) + \theta l_{n,21}^{H_1}(1, 1, \hat{\eta}) - (L_{13}^T(\hat{\eta}) + \theta L_{23}^T(\hat{\eta})) L_{33}^{-1}(\hat{\eta}) l_{n,31}^{H_1}(1, 1, \hat{\eta}))$$

step 8 Compute $e_P(Z_{MERT}, Z(\theta^{(0)}))$.

$$e_P(Z_{MERT}, Z(\theta^{(0)})) = \frac{(\rho_{\theta_i, \theta^{(0)}} + \rho_{\theta_j, \theta^{(0)}})^2}{2(1 + \rho_{\theta_i, \theta_j})}$$

$$\rho_{\theta_i, \theta_j} = \frac{\sigma(\hat{\eta}, \theta_i, \theta_j)}{\sigma(\hat{\eta}, \theta_i, \theta_i)\sigma(\hat{\eta}, \theta_i, \theta_j)^{1/2}}$$

step 9 Compute

$$\begin{aligned}\tilde{e}_C(Z_{MERT}, Z_{\theta^{(0)}}) &= \frac{\tilde{Q}_{Z_{MERT}}}{\tilde{Q}_{Z_{\theta^{(0)}}}} \\ \tilde{Q}_{Z_{\theta^{(0)}}} &= 2 \left(1 - \Phi \left(\frac{\mu^{(1)}(\lambda_0, \theta^{(0)})}{2\sigma(\theta^{(0)})} \right) \right) \\ \tilde{Q}_{Z_{MERT}} &= 2 \left(1 - \Phi \left(\left[\frac{\mu^{(1)}(\lambda_0, \theta_i)}{2\sigma(\theta_i)} + \frac{\mu^{(1)}(\lambda_0, \theta_j)}{2\sigma(\theta_j)} \right] / \sqrt{8(1 + \rho_{\theta_i, \theta_j})} \right) \right) \\ \lambda_0 &= 1\end{aligned}$$

step 10 For each λ . Compute

$$\begin{aligned}e_{HL}(Z_{MERT}, Z_\theta) &= \frac{d_{Z_{MERT}}(\lambda)}{d_{Z_\theta}(\lambda)} \\ d_{Z_\theta}(\lambda) &= \frac{\mu^2(\lambda, \theta)}{\sigma^2(\theta)} \\ d_{Z_{MERT}}(\lambda) &= \mu_{MERT}^2(\lambda) \\ \mu_{MERT}(\lambda) &= [\mu(\lambda, \theta_i)/\sigma(\theta_i) + \mu(\lambda, \theta_j)/\sigma(\theta_j)] / \sqrt{2(1 + \rho_{\theta_i, \theta_j})}\end{aligned}$$

step 11 Compute

$$\begin{aligned}e_B(Z_{MERT}, Z_\theta) &= e_{HL}(Z_{MERT}, Z_\theta) \\ \tilde{e}_B(Z_{MERT}, Z_\theta) &= \frac{\tilde{c}_{Z_{MERT}}}{\tilde{c}_{Z_\theta}} \\ \tilde{c}_{Z_{MERT}} &= 1 - \Phi(\mu_{MERT}^{(1)}(1)) \\ \tilde{c}_{Z_\theta} &= 1 - \Phi(\mu^{(1)}(1, \theta^{(0)})/\sigma(\theta^{(0)})) \\ \mu_{MERT}^{(1)}(\lambda) &= [\mu^{(1)}(\lambda, \theta_i)/\sigma(\theta_i) + \mu^{(1)}(\lambda, \theta_j)/\sigma(\theta_j)] / \sqrt{2(1 + \rho_{\theta_i, \theta_j})}\end{aligned}$$

1.3 Simulation

Let p be the minor allele frequency (MAF) of the marker of interest in the population. we consider case-control data with $r = 500$ cases and $s = 500$ controls without covariates. and $\lambda \in \{1.1, 1.3, 1.5\}$ and $p \in \{0.15, 0.30, 0.45\}$, and the true $\theta^{(0)} \in \{0, 1/4, 1/2, 1\}$, and the disease prevalence $k = 0.05$. we generate $Nrep = 1000$ datasets. and we compute the mean and variance of the four AREs to Z_{MERT} with $\theta_i = 0, \theta_j = 1$ and $Z_{\theta^{(0)}}$

simulate case-control data[(Gang Zheng, Yaning Yang, Xiaofeng Zhu, Robert C. Elston Analysis of Genetic Association Studies (2012)] To simulate case-control samples without covariates for genetic marker M given a specific genetic model, the following algorithm can be used:

1. Specify the numbers of cases (r) and controls (s), the disease prevalence k , the allele frequency p for the risk allele B (the minor allele if the risk allele is unknown), and the GRR $\lambda_2 = \lambda$;
2. Calculate the GRR λ_1 based on the given genetic model and population genotype frequencies g_0, g_1 and g_2 under Hardy-Weinberg proportions in the population;

Table 1: The Four AREs of Z_{MERT} and $Z_{\theta^{(0)}}$

MAF	$\theta^{(0)}$	$\lambda = 1.1$				$\lambda = 1.3$				$\lambda = 1.5$			
		e_P	e_C	e_{HL}	e_B	e_P	e_C	e_{HL}	e_B	e_P	e_C	e_{HL}	e_B
0.15	0	0.974 (0.0069)		0.872 (0.017)									
	1/4												
	1/2												
	1												
0.30	0												
	1/4												
	1/2												
	1												
0.45	0												
	1/4												
	1/2												
	1												

3. Calculate $f_0 = k/(g_0 + \lambda_1 g_1 + \lambda_2 g_2)$, $f_1 = \lambda_1 f_0$, and $f_2 = \lambda_2 f_0$;
4. Calculate $p_j = g_j f_j / k$ and $q_j = g_j(1 - f_j)/(1 - k)$ for $j = 0, 1, 2$;
5. Generate random samples (r_0, r_1, r_2) and (s_0, s_1, s_2) independently from the multinomial distributions $Mul(r; p_0, p_1, p_2)$ and $Mul(s; q_0, q_1, q_2)$, respectively