File S1

Approximate equivalence between BIMBAM and CAVIAR for binary traits

For binary traits, we use a logistic model as follows:

$$log\frac{p(y_i=1)}{p(y_i=0)} = \alpha + \sum_{j=1}^{m} X_{ij}^T\beta_j\,, i = 1, \cdots, n \qquad \text{(A1)}$$

where y_i is the phenotype of individual i; 1 indicates a case and 0 indicates a control. X_{ij} is the same additively coded genotype of individual i and SNP j as for the above quantitative traits. The model parameters are α and $\beta = (\beta_1, \cdots, \beta_m)^T$. The phenotype vector is $y = (y_1, \dots, y_n)^T$. Again we assume each column of X has mean 0 and variance 1, i.e., $\frac{1}{n}\sum_{i=1}^{n} X_{ij} = 0, \frac{1}{n}\sum_{i=1}^{n} X_{ij}^2 = 1, j = 1, 2, \dots, n$. Denote the number of cases and controls by n_1 and n_2, respectively. Let $\bar{X} = (1_{n\times1}, X), \bar{\beta} = (\alpha, \beta)^T$, where $1_{n\times1}$means the $n \times 1$ vector of 1s. We assume a normal prior distribution for $\bar{\beta}$, i.e., $\bar{\beta} \sim N(0, \bar{v})$, where $\bar{v}$ is a diagonal matrix with positive diagonal entries. Denote the first diagonal entry for α by σ_α^2, the rest of the diagonal matrix by v, the variance of β. The null model is $\beta = 0_{m\times1}$, which is equivalent to setting v to $0_{m\times m}$, where $0_{m\times1}$ is a $m \times 1$ vector of 0s, and $0_{m\times m}$ is a $m \times m$ matrix of 0s. The Bayes factor comparing the full model with the null model is

$$BF = \frac{p(y, \bar{X}|\sigma_\alpha^2, v)}{p(y, \bar{X}|\sigma_\alpha^2, v = 0_{m\times m})},$$

where $p(y, \bar{X}|\sigma_\alpha^2, v) = \int p(y, \bar{X}|\bar{\beta})p(\bar{\beta}|\sigma_\alpha^2, v)d\bar{\beta}$, an integral over the prior distribution of $\bar{\beta}$. BIMBAM approximates the integral using Laplace's method. In the following, we approximate the integral using sufficient statistics and normal distributions.

For canonical link functions of generalized linear models, $(\bar{X}^T y, \bar{X})$ are the sufficient statistics for $\bar{\beta}$ (AGRESTI 2013). We also consider $\bar{X}$ as random. By the definition of sufficient statistics

$$p(y, \bar{X}|\bar{\beta}) = p(\bar{X}^T y\,, \bar{X}|\bar{\beta})p(y, \bar{X}|\bar{X}^T y\,, \bar{X}) = p(\bar{X}^T y\,|\bar{X}, \bar{\beta})p(\bar{X}|\bar{\beta})p(y, \bar{X}|\bar{X}^T y\,, \bar{X}),$$

where $p(\bar{X}^T y\,,\bar{X}|\bar{\beta})$ is the likelihood of $(\bar{X}^T y\,,\bar{X})$ given $\bar{\beta}$, $p(y,\bar{X}|\bar{X}^T y\,,\bar{X})$ is the conditional probability of the data $(y,\bar{X})$ given $(\bar{X}^T y\,,\bar{X})$, which does not depend on $\bar{\beta}$. Because $\bar{X}$ does not depend on $\bar{\beta}$, we have

$$p(y,\bar{X}|\bar{\beta}) = p(\bar{X}^T y\,|\bar{X},\bar{\beta})p(\bar{X})p(y,\bar{X}|\bar{X}^T y\,,\bar{X}).$$

Therefore the Bayes factor can be written as the ratio of two likelihoods

$$BF = \frac{\int p(\bar{X}^T y\,|\bar{X},\bar{\beta})p(\bar{\beta}|\sigma_\alpha^2,v)d\bar{\beta}}{\int p(\bar{X}^T y\,|\bar{X},\bar{\beta})p(\bar{\beta}|\sigma_\alpha^2,v=0)d\bar{\beta}}.$$

Denote the numerator by L_1 and the denominator by L_0. They are the marginal likelihood of $\bar{X}^T y$ given $\bar{X}$ after integrating out $\bar{\beta}$. Now we approximate this marginal likelihood using normal distributions. From the Central Limit Theorem, $\bar{X}^T y$ given $\bar{X}$ and $\bar{\beta}$ has an approximate multivariate normal distribution with the following mean and variance

$$E(\bar{X}^T y\,|\bar{X},\bar{\beta}) = \bar{X}^T E(y) = \bar{X}^T P,$$

$$Var(\bar{X}^T y\,|\bar{X},\bar{\beta}) = \bar{X}^T W \bar{X}$$

where P is a vector with each element $p_i(\alpha,\beta) = p(y_i = 1) = 1/\,(1+\exp(-(\alpha + \sum_{j=1}^{m} X_{ij}^T \beta_j)))$, W is a $n \times n$ diagonal matrix with the ith diagonal entry $p_i \times (1-p_i)$. Because the effects β are usually small in real data, we can approximate W by W_0, the estimated variance under the null hypothesis. Specifically, $W_0 = \tilde{y}_i(1-\tilde{y}_i)I_n = w_0 I_n$, where $w_0 = \frac{n_1 n_2}{n^2}$. We can also linearize p_i, the mean of y_i, using the Taylor expansion at the MLE of the null hypothesis, denoted by α_0 and $\beta_0 = 0_{m\times 1}$, where $p_i(\alpha_0,\beta_0) = \tilde{y}_i = \frac{n_1}{n}$. Specifically,

$$p_i(\alpha,\beta) \approx p_i(\alpha_0,\beta_0) + p_i(\alpha_0,\beta_0)\big(1 - p_i(\alpha_0,\beta_0)\big)\left((\alpha - \alpha_0) + \sum_{j=1}^{m} X_{ij}^T\beta_j\right)$$

$$= \frac{n_1}{n} + \frac{n_1n_2}{n^2}(-\alpha_0) + \frac{n_1n_2}{n^2}\sum_{j=1}^{m}(\alpha + X_{ij}^T\beta_j).$$

In the matrix form, $P = t_1 1_{n\times 1} + w_0\bar{X}\bar{\beta}$, where $t_1 = \frac{n_1}{n} + \frac{n_1n_2}{n^2}(-\alpha_0)$, $1_{n\times 1}$ is a $n \times 1$ vector of 1s. In summary, we can write the approximate normal distribution of $\bar{X}^Ty$ given $\bar{X}$ and $\bar{\beta}$ as follows:

$$E\big(\bar{X}^Ty\,|\bar{X},\bar{\beta}\big) \approx t_1\bar{X}^T1_{n\times 1} + w_0\bar{X}^T\bar{X}\bar{\beta},$$

$$Var\big(\bar{X}^Ty\,\big|\bar{X},\bar{\beta}\big) \approx w_0\bar{X}^T\bar{X}$$

Because $\bar{\beta}$ follows a multivariate normal distribution $\bar{\beta}\sim N(0,\bar{v})$, the marginal distribution of $\bar{X}^Ty$ given $\bar{X}$ also has a multivariate normal distribution. Specifically, we have

$$E(\bar{X}^Ty\,|\bar{X}) = E\left(E\big(\bar{X}^Ty\,|\bar{X},\bar{\beta}\big)\right) \approx \begin{pmatrix} nt_1 \\ 0_{m\times 1}\end{pmatrix},$$

$$Var(\bar{X}^Ty\,|\bar{X}) = E\left(Var\big(\bar{X}^Ty\,|\bar{X},\bar{\beta}\big)\right) + Var\left(E\big(\bar{X}^Ty\,|\bar{X},\bar{\beta}\big)\right)$$

$$\approx w_0\bar{X}^T\bar{X} + w_0^2\bar{X}^T\bar{X}\bar{v}\bar{X}^T\bar{X} = \begin{pmatrix} nw_0 + n^2w_0^2\sigma_\alpha^2 & \\ & w_0X^TX + w_0^2X^TXvX^TX\end{pmatrix}$$

Therefore likelihood L_1 can be approximated as

$$\hat{L}_1 = (2\pi)^{-\frac{m+1}{2}}\left(\left|nw_0 + n^2w_0^2\sigma_\alpha^2\right|\left|w_0X^TX\right|\left|I_m + w_0vX^TX\right|\right)^{-\frac{1}{2}}\exp\left(-\frac{1}{2}D_1\right),$$

where

$$D_1 = (1_{nx1}^Ty - nt_1)^T(nw_0 + n^2w_0^2\sigma_\alpha^2)^{-1}(1_{nx1}^Ty - nt_1) +$$

$$(X^Ty)^T(w_0X^TX + w_0^2X^TXvX^TX)^{-1}(X^Ty).$$

By setting v to $0_{m\times m}$, we get the approximated L_0 as

$$\hat{L}_0 = (2\pi)^{-\frac{m+1}{2}}\left(\left|nw_0 + n^2 w_0^2 \sigma_\alpha^2\right| |w_0 X^T X|\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2} D_0\right),$$

where $D_0 = (1_{nx1}^T y - nt_1)^T (nw_0 + n^2 w_0^2 \sigma_\alpha^2)^{-1} (1_{nx1}^T y - nt_1) + (X^T y)^T (w_0 X^T X)^{-1} (X^T y)$. From Woodbury matrix identity,

$$(w_0 X^T X + w_0^2 X^T X v X^T X)^{-1} = (w_0 X^T X)^{-1} - (v^{-1} + w_0 X^T X)^{-1}.$$

Therefore the approximate Bayes factor is

$$\widehat{BF} = |I_m + w_0 v X^T X|^{-\frac{1}{2}} \exp\left(\frac{1}{2} y^T X (v^{-1} + w_0 X^T X)^{-1} X^T y\right).$$

By plugging in the Armitage trend test statistic $z = \sqrt{\frac{n}{n_1 n_2}} X^T y$ (see the derivation from the following section about non-centrality parameters), $\Sigma_x = \frac{X^T X}{n}$ and $w_0 = \frac{n_1 n_2}{n^2}$,

$$\widehat{BF} = |I_m + n w_0 v \Sigma_x| \exp\left(\frac{1}{2} z^T ((n w_0 v)^{-1} + \Sigma_x)^{-1} z\right).$$

This is the same as equation (3) except the coefficient w_0, therefore completing the proof.

We also note that here v is the variance of β, while in the proof for quantitative traits, the variance of β is $v\frac{1}{\tau}$. Let $v = \sigma_a^2 I_m$. The input for BIMBAM, denoted by $\sigma_a(BIMBAM)$, is σ_a, while the input for CAVIARBF, denoted by $\sigma_a(CAVIARBF)$, is $\sqrt{w_0}\sigma_a$. To get results similar to BIMBAM with the "-cc" option, in addition to setting the weights to the variances of SNPs as in quantitative traits, we also need to make sure that

$$\sigma_a(CAVIARBF) = \sqrt{w_0}\sigma_a(BIMBAM) = \sqrt{\frac{n_1 n_2}{n^2}}\ \sigma_a(BIMBAM).$$

Non-centrality parameters of the marginal test statistics under multiple causal SNPs

For quantitative traits, without loss of generality, we can assume the same model as in equation (1). We rewrite it here and use σ^2 instead of $\frac{1}{\tau}$:

$$y = X\beta + \ \varepsilon, \varepsilon \sim N(0, \sigma^2 I_n).$$

Each column of X has mean 0 and variance 1, i.e., $\frac{1}{n}\sum_{i=1}^{n} X_{ij} = 0, \frac{1}{n}\sum_{i=1}^{n} X_{ij}^2 = 1, j = 1, 2, \dots, n$. Denote the column j of X by X_j, so that the marginal test statistic is

$$z_j = \frac{\left(X_j^T X_j\right)^{-1} X_j^T y}{\hat{\sigma}_j \left(X_j^T X_j\right)^{-\frac{1}{2}}} = \frac{\left(X_j^T X_j\right)^{-\frac{1}{2}} X_j^T y}{\hat{\sigma}_j} = \frac{n^{-\frac{1}{2}} X_j^T y}{\hat{\sigma}_j}.$$

Assume $\hat{\sigma}_j$ is a good approximation of σ when the sample size is large enough. This assumption is acceptable because the proportion of variation explained by X is usually small. Therefore the test statistic can be approximated by $\hat{z}_j = \frac{n^{-\frac{1}{2}} X_j^T y}{\sigma}$. Let $\hat{z} = [\hat{z}_1, \hat{z}_2, \cdots, \hat{z}_m]^T$. In matrix form, we have

$$\hat{z} = \frac{n^{-\frac{1}{2}} X^T y}{\sigma}.$$

Therefore,

$$E(\hat{z}) = \frac{n^{-\frac{1}{2}} X^T X \beta}{\sigma} = \frac{n^{\frac{1}{2}} \Sigma_x \beta}{\sigma},$$

$$Var(\hat{z}) = \frac{1}{n\sigma^2} X^T var(y) X = \frac{X^T X}{n} = \Sigma_x,$$

where $\Sigma_x = \frac{X^T X}{n}$. This also shows the approximate multivariate normal distribution for the marginal test statistics. The marginal non-centrality parameter for each SNP is the square of each element in $E(\hat{z})$. With the marginal non-centrality parameters, we can calculate the power for the causal SNPs.

For binary traits, we use the model specified in equation A1. For simplicity, we first assume data are generated in a prospective logistic model. Following (SCHAID *et al.* 2002; SEAMAN and MULLER-MYHSOK 2005), the score statistic vector for each SNP is

$$U_\beta = \left(U_{\beta_1}, \cdots, U_{\beta_m}\right)^T, U_{\beta_j} = \sum_{i=1}^{n} (y_i - \tilde{y}_i) X_{ij}\,, j = 1, \cdots, m,$$

where $\tilde{y}_i$ is the fitted value for individual i, which is obtained under the null hypothesis, i.e., setting all $\beta_j, j = 1, \cdots, m$ to 0, to obtain the maximum likelihood estimate $\hat{\alpha}$ of α and then calculate the fitted $\tilde{y}_i$. Under the null hypothesis that $\beta = 0$, the variance of U_β is

$$V_\beta = \tilde{y}(1 - \tilde{y})(X^T X - n x_m x_m^T),$$

where $\tilde{y} = (\tilde{y}_1, \cdots, \tilde{y}_n)^T$, x_m is a column vector where each element is the mean of each column in matrix X. Under the null hypothesis, U_β is asymptotically distributed multivariate normal, i.e., $U_\beta \sim N\left(0, V_\beta\right)$ or $U_\beta^T V_\beta^{-1} U_\beta$ has a chi-square distribution. Because there are no other covariates except the intercept and X is centered and scaled, we have $\tilde{y}_i = \frac{n_1}{n}$, $U_\beta = X^T y$, and $V_\beta = \frac{n_1 n_2}{n^2} X^T X = \frac{n_1 n_2}{n} \Sigma_x$, where $\Sigma_x = \frac{X^T X}{n}$. The marginal score test statistic for SNP j can be obtained by only keeping the jth column in X in the model. Specifically, the marginal score test statistic vector is

$$z = \frac{U_\beta}{\sqrt{\frac{n_1 n_2}{n}}} = \sqrt{\frac{n}{n_1 n_2}} X^T y = \frac{n^{-\frac{1}{2}} X^T y}{\hat{\sigma}},$$

where $\hat{\sigma} = \sqrt{\frac{n_1 n_2}{n^2}} = \sqrt{\tilde{y}_i(1 - \tilde{y}_i)}$, the estimated standard deviation of y. The test statistics have a similar form as that for quantitative traits. These are also Armitage's trend tests (SASIENI 1997).

To calculate the power, we need to know the distribution under the alternative hypothesis. When the sample size is large, based on the Central Limit Theorem, z has a multivariate normal distribution. We have

$$E(z) = \sqrt{\frac{n}{n_1 n_2}} X^T E(y) = \sqrt{\frac{n}{n_1 n_2}} X^T P,$$

$$Var(z) = \frac{n}{n_1 n_2} X^T var(y) X = \frac{n}{n_1 n_2} X^T W X,$$

where P is a vector with each element $p_i(\alpha, \beta) = p(y_i = 1) = 1/\,(1 + \exp(-(\alpha + \sum_{j=1}^{m} X_{ij}^T \beta_j)))$, W is a $n \times n$ diagonal matrix with the ith diagonal entry $p_i \times (1 - p_i)$. With known α and β, the power of Armitage's trend test can be calculated. For retrospective case control studies, we should change α to α^* to reflect the different sampling probabilities for cases and controls (AGRESTI 2013). Because the effects β are usually small in real data, we can approximate W by W_0, the estimated variance under the null hypothesis. Specifically, $W_0 = \tilde{y}_i(1 - \tilde{y}_i) I_n$. Therefore we have

$$Var(z) \approx \frac{X^T X}{n} = \Sigma_x.$$

We can also linearize p_i, the mean of y_i, using the Taylor expansion at the MLE of the null hypothesis, denoted by α_0 and $\beta_0 = 0_{m \times 1}$, where $p_i(\alpha_0, \beta_0) = \tilde{y}_i = \frac{n_1}{n}$. Therefore,

$$p_i(\alpha, \beta) \approx \frac{n_1}{n} + \frac{n_1 n_2}{n^2}(\alpha - \alpha_0) + \frac{n_1 n_2}{n^2} \sum_{j=1}^{m} X_{ij}^T \beta_j.$$

In the matrix form, $P = t_2 1_{n \times 1} + \hat{\sigma}^2 X \beta$, where $t_2 = \frac{n_1}{n} + \frac{n_1 n_2}{n^2}(\alpha - \alpha_0)$. Therefore

$$E(z) \approx \sqrt{\frac{n}{n_1 n_2}} X^T \frac{n_1 n_2}{n^2} X \beta = \sqrt{\frac{n_1 n_2}{n}} \Sigma_x \beta = n^{\frac{1}{2}} \hat{\sigma} \Sigma_x \beta.$$

The marginal non-centrality parameter for each SNP is the square of each element in $E(z)$. In this approximation the non-centrality parameters do not require the specification of the intercept. This also proves the approximate multivariate normal distribution of the marginal test statistics under the logistic model. We can see that the approximate distributions of the marginal test statistics have a similar form as quantitative traits.

LITERATURE CITED

AGRESTI, A., 2013 *Categorical data analysis*. Wiley, Hoboken, NJ.

SASIENI, P. D., 1997 From genotypes to genes: doubling the sample size. Biometrics **53:** 1253-1261.

SCHAID, D. J., C. M. ROWLAND, D. E. TINES, R. M. JACOBSON and G. A. POLAND, 2002 Score tests for association between traits and haplotypes when linkage phase is ambiguous. American journal of human genetics **70:** 425-434.

SEAMAN, S. R., and B. MULLER-MYHSOK, 2005 Rapid simulation of P values for product methods and multiple-testing adjustment in association studies. American journal of human genetics **76:** 399-408.