Hindawi Publishing Corporation
Journal of Probability and Statistics
Volume 2012, Article ID 419832, 12 pages
doi:10.1155/2012/419832

Research Article

A Multinomial Ordinal Probit Model with Singular Value Decomposition Method for a Multinomial Trait

Soonil Kwon,[1] Mark O. Goodarzi,[1] Kent D. Taylor,[1] Jinrui Cui,[1] Y.-D. Ida Chen,[1] Jerome I. Rotter,[1] Willa Hsueh,[2] and Xiuqing Guo[1]

[1] *Medical Genetics Institute, Cedars-Sinai Medical Center, 8700 Beverly Boulevard, Paciffc Theatres Building, 4th Floor, Los Angeles, CA 90048, USA*
[2] *The Methodist Hospital Research Institute, The Methodist Hospital, 6670 Bertner Street R8-103, Houston, TX 77030, USA*

Correspondence should be addressed to Xiuqing Guo, xiuqing.guo@cshs.org

Received 12 March 2012; Accepted 31 March 2012

Academic Editor: Yongzhao Shao

We developed a multinomial ordinal probit model with singular value decomposition for testing a large number of single nucleotide polymorphisms (SNPs) simultaneously for association with multidisease status when sample size is much smaller than the number of SNPs. The validity and performance of the method was evaluated via simulation. We applied the method to our real study sample recruited through the Mexican-American Coronary Artery Disease study. We found 3 genes (SORCS1, AMPD1, and PPARα) to be associated with the development of both IGT and IFG, while 5 genes (AMPD2, PRKAA2, C5, TCF7L2, and ITR) with the IGT mechanism only and 6 genes (CAPN10, IL4, NOS3, CD14, GCG, and SORT1) with the IFG mechanism only. These data suggest that IGT and IFG may indicate different physiological mechanism to prediabetes, via different genetic determinants.

1. Introduction

Genome-wide association studies (GWASs) examine genetic variants across the entire genome to improve the understanding of genetic components underlying complex human disease. With whole-genome genotyping techniques that allow GWAS to involve hundreds of thousands of single nucleotide polymorphisms (SNPs), many studies have successfully identified novel genetic components for many diseases or related quantitative traits. However, the sample size is often limited due to the difficulty of recruiting patients and/or the cost of research. This leads to the situation that the number of SNPs (m) is much larger than the samples (n) available, that is, $m \gg n$, and makes the traditional statistical methods unsuitable for analyzing multiple SNPs simultaneously. Most of current GWASs deal with this shortcoming by performing single SNP association test, which analyzes one SNP at a time and results in a huge multiple testing problem. These motivated us to develop methods that can avoid the multiple testing problem, in other words, methods that can evaluate multiple SNPs simultaneously when $m \gg n$. We first introduced the iterative Bayesian variable selection (IBVS) method [1], which analyzes all SNPs simultaneously when $m \gg n$ and uses the Bayesian variable selection [2] iteratively to find SNPs that are associated with disease. The method was successfully applied to the simulated rheumatoid arthritis data provided by the Genetic Analysis Workshop 15 (GAW15). We later introduced the Bayesian classification with singular value decomposition (BCSVD) method [3]. The method applies the singular value decomposition (SVD) to the covariate matrix, which is usually the genotype data in GWAS and reduces the dimension of parameters to be estimated to the number of samples. This makes the method feasible to handle multiple SNPs simultaneously when $m \gg n$. The validation of the method was demonstrated by applying to the simulated data provided by GAW16. We now extend our method from binary disease to the analysis of polytomous ordinal response variables. We propose here a multinomial ordinal probit model with singular value decomposition method. We show the validity of the newly developed method by applying it to simulated data sets as well as to a real study sample to identify genes contributing to two different mechanisms for prediabetes, namely, impaired glucose tolerance (IGT) and impaired fasting glucose (IFG). With the simulated data sets, we demonstrate that this new method is superior to single SNP analysis method and, with the real data, identify different genes for each mechanism.

2. Method and Materials

2.1. Multinomial Probit Model with Singular Value Decomposition

Logit and probit models are statistical models that are widely used for the analysis of categorical (ordinal/nominal) data. The difference between these two models is the choice of the link function relating the linear predictor to the expected value; the probit model uses the inverse normal cumulative distribution, and the logit model uses the logit transformation. As discussed by Greene [4], in most cases, the choice of the link function is largely a matter of taste. We utilized the probit model here to analyze data with polytomous ordinal response variables. In general, the multinomial ordinal probit model can be expressed by latent (unobserved) continuous variables associated with categorical responses. Let us assume that responses $y_1, y_2, \dots, y_n$ are observed, where y_i takes one of the J-ordered categories and $\theta_1, \dots, \theta_J$ are real numbers of bin boundaries, which satisfy that $-\infty = \theta_1 \leq \cdots \leq \theta_J = \infty$. As discussed by Albert and Chib [5], we denote that $z_1, z_2, \dots, z_n$ are latent continuous random variables. We assume that the latent variable (z_i) associated with a categorical outcome (y_i) can be explained in terms of an underlying linear model, and that the observed response y_i has category j if and only if z_i falls between θ_{j-1} and θ_j. The multinomial ordinal probit model is equivalent to the following model:

$$\begin{gathered} z_i = x_i\beta + \epsilon_i, \quad \epsilon_i \sim N(0,1), \quad i = 1, \dots, n, \\ y_i = j \iff \theta_{j-1} < z_i \leq \theta_j, \quad j = 1, \dots, J, \end{gathered} \tag{2.1}$$

where x_i is a $1 \times m$ vector of the explanatory variables for the ith sample, and β is a $m \times 1$ vector of parameters to be estimated. In vector-matrix notation, we can have the multinomial ordinal probit model

$$z = X\beta + \epsilon, \tag{2.2}$$

where z is the $n\times 1$ vector of latent variables, X is the $n\times m$ matrix of the explanatory variables, β is the $m \times 1$ vector of unknown regression coefficients, and ϵ follows an independent standard multivariate normal distribution, $\epsilon \sim N(0, I_n)$. By applying SVD to the matrix X in (2.2), when $\text{rank}(X) = n$, the matrix can be expressed as $X' = ADF'$, where A is the $m \times n$ singular value factor loading matrix with orthonormal columns so that $A'A = I_n$, F is the $n \times n$ SVD orthogonal factor matrix with $F'F = FF' = I_n$, and $D = \text{diag}(d_1, \dots, d_n)$, the diagonal matrix of positive singular values, ordered as $d_1 \geq \cdots \geq d_n > 0$. When $\text{rank}(X) = r < n$, the smallest $n - r$ singular values in D are replaced with 0, that is, $d_1 \geq \cdots \geq d_r > d_{r+1} = \cdots = d_n = 0$. Therefore, in the product $X' = ADF'$, the last $n - r$ columns of both A and F for which $d_{r+1} = \cdots = d_n = 0$ are ignored since they interact with the block of zeros in D. Hence, this leads to another form of SVD, $X' = A_r D_r F_r'$, that is, the product of the first r columns of A, the upper $r \times r$ block of D, and the first r columns of F. Since the difference between the both scenario is only in dimension of matrices in SVD, we assume that $\text{rank}(X) = n$ in the rest part of the paper for convenience. Thus, the model in (2.2) with the SVD of X can be written as follows:

$$z = X\beta + \epsilon = \left(ADF'\right)'\beta + \epsilon = FDA'\beta + \epsilon = L\gamma + \epsilon, \tag{2.3}$$

where $L = FD$ and $\gamma_{n\times 1} = A'_{\ n\times m}\beta_{m\times 1}$. Therefore, z, the $n \times 1$ vector of latent variables in (2.3), has a multivariate normal distribution, that is, $z \sim N(L\gamma, I_n)$. As shown in (2.3), γ is expressed by a linear combination of the original parameters (β). Hence, we call γ as the vector of superfactors. The model in (2.3) represents a massive dimension reduction from m to n parameters. The regression model with m parameters reduced to that with n parameters derived from the SVD of the covariate matrix X. Therefore, the statistical inference on the original parameter turns into the superfactors. Let $p_i = (p_{i1}, \dots, p_{iJ})$ denote the vector of probabilities associated with the assignment of the ith sample into categories $1, \dots, J$, where p_{ij} denote the probability that a sample falls into category j. From (2.1) and (2.3), it follows that

$$p_{ij} = \int_{\theta_{j-1}}^{\theta_j} \phi(z - l_i\gamma)dz = \Pr(\theta_{j-1} < Z_i < \theta_j) = \Phi(\theta_j - l_i\gamma) - \Phi(\theta_{j-1} - l_i\gamma), \tag{2.4}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the probability density function and the cumulative density function of the standard normal distribution, respectively, and l_i is the ith row of matrix $L = FD$. Let $y = (y_1, \dots, y_n)$ denote the vector of responses observed for all samples. Then, the probability of observing data y is given as follow:

$$\Pr(y \mid p_i) = \prod_{i=1}^{n}\prod_{j=1}^{J} p_{ij}^{I\{y_i = j\}}. \tag{2.5}$$

From (2.4) and (2.5), the log likelihood function for (γ, θ) can be written as

$$\ln L(\gamma, \theta) = \sum_{i=1}^{n} \sum_{j=1}^{J} I\{y_i = j\} \ln\left[\Phi(\theta_j - l_i\gamma) - \Phi(\theta_{j-1} - l_i\gamma)\right]. \tag{2.6}$$

2.2. Model Fitting with Maximum Likelihood Estimation

The maximum likelihood estimates (MLEs) of the superfactors, γ, in (2.3) can be obtained by the iteratively reweighted least squares (IRLSs) procedure [6] using the log likelihood function for (γ, θ) in (2.6). The procedure can be briefly described as follows. Let η denote the vector of all model parameters, that is, $\eta = (\theta_2, \ldots, \theta_{J-1}, \gamma_1, \ldots, \gamma_{n-J+2})$. Note that θ_1 and θ_J are not included in this vector because their values are assumed to be 0 and ∞, respectively, for the purpose of model identifiability. Also note that the $(J - 2)$ smallest singular values together with their corresponding factors are dropped from the parameters since the number of parameters must not exceed the number of samples. Assuming that $J = 4$, define that

$$\mathcal{C}_i = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{bmatrix}, \qquad \mathcal{L}_i = \begin{bmatrix} 1 & 0 & 0 & -l_i \\ 0 & 1 & 0 & -l_i \\ 0 & 0 & 1 & -l_i \end{bmatrix}, \tag{2.7}$$

and $\mathcal{H}_i = \operatorname{diag}(f_{i1}, \ldots, f_{iJ-1})$, where f_{ij} denotes the derivative of the standard normal cumulative distribution function at $\theta_j - l_i\gamma$. Take $W_i = \operatorname{diag}(p_i)$, where p_i is the $J \times 1$ vector of probabilities that the ith individual falls in each category, that is, $p_i = (p_{i1}, \ldots, p_{i_J})'$, and let $\mathcal{N}_i$ be a $J \times 1$ vector of observation, that is, $\mathcal{N}_i = (I\{y_i = 1\}, \ldots, I\{y_i = J\})'$. After initialization of all elements, the iteration $s + 1$ $(s = 1, 2, \ldots)$ can be written as

$$\eta^{s+1} = \eta^s + \left(\sum_{i=1}^{n} \mathcal{L}_i' \mathcal{H}_i^{(s)} \mathcal{C}_i' W_i^{-1(s)} \mathcal{C}_i \mathcal{H}_i^{(s)} \mathcal{L}_i \right)^{-1} \sum_{i=1}^{n} \mathcal{L}_i' \mathcal{H}_i^{(s)} \mathcal{C}_i' W_i^{-1(s)} \left(\mathcal{N}_i - p_i^{(s)} \right). \tag{2.8}$$

The MLE of η can be found by performing the process recursively until the change between η^{s+1} and η^s is negligible.

2.3. General Solution for the Original Parameters

We have discussed how to estimate the superfactor (γ) in (2.3) thus far. Since the primary interest is to find SNPs that are significantly associated with a disease, it is necessary to transform the superfactor (γ) to the original parameters (β) in (2.1). The equation $\gamma = A'\beta$ in (2.3) can be utilized for the transformation even though A is $m \times n$ nonsquare matrix. As discussed by Graybill [7], the unique solution for β can be achieved by taking the generalized inverse matrix of A' as A since $A'A = I_n$. Therefore, the unique solution for SNP effect (β) can be calculated by $\beta = A\gamma$.

2.4. Selection of Significant SNPs

Finding significant SNPs is the same as testing if each SNP effect (β_i, $i = 1, \ldots, m$) is statistically significant, that is, testing the hypothesis: H_0: $\beta_i = 0$ versus H_1: $\beta_i \neq 0, i = 1, \ldots, m$. The simple method is to use Wald's test statistic, which forms $(\widehat{\beta} - \beta)/\text{se}(\widehat{\beta})$ and assumes a normal distribution. However, when $m \gg n$, it is hard to calculate $\text{se}(\widehat{\beta})$ directly from the data. We therefore utilized permutation test to select significant SNPs. The rationale behind the test is that, under the null hypothesis, the estimate of β obtained from the raw (unpermuted) data is similar to the estimate of β obtained from the permuted data. That is, the difference between two estimates is closed to zero under H_0. With this idea, we can construct Wald's test statistic as follows. Let $\widehat{\beta}_i$ $(i = 1, \ldots, m)$ be the estimate of the ith SNP effect from the raw data and $\widehat{\beta}_i^k$ $(k = 1, \ldots, K)$ be the estimate of the ith SNP effect from the kth-permuted data. Let us define $\beta_i^{d_k}$ as the difference between $\widehat{\beta}_i$ and $\widehat{\beta}_i^k$, that is, $\beta_i^{d_k} = \widehat{\beta}_i - \widehat{\beta}_i^k$. Then, Wald's test statistic can be as follows:

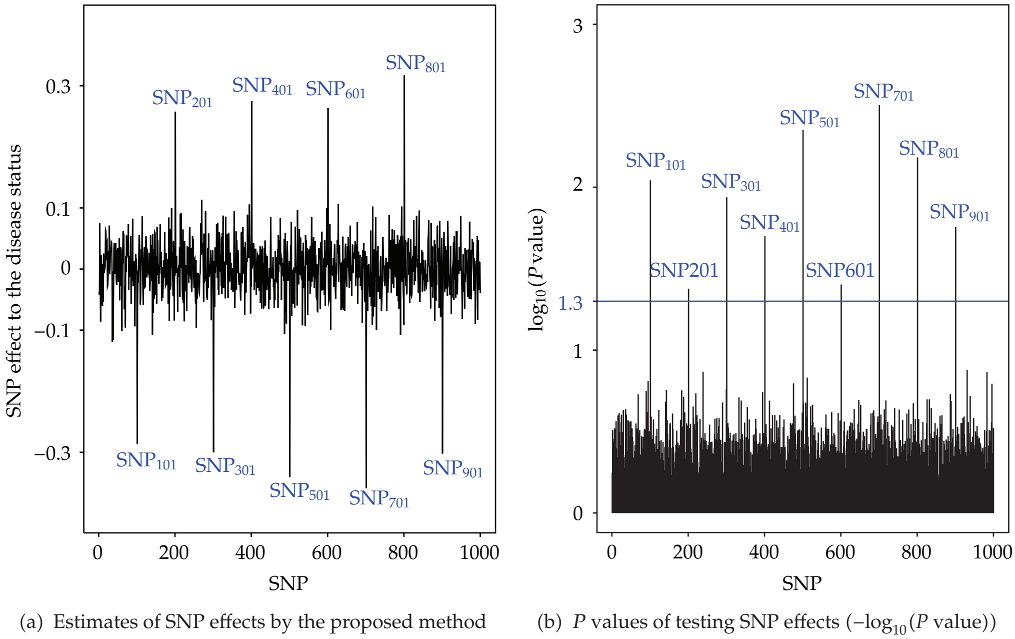
$$\Lambda_i = \frac{\overline{\beta}_i^d}{\text{se}\left(\overline{\beta}_i^d\right)}, \quad i = 1, \ldots, m, \tag{2.9}$$

where $\overline{\beta}_i^d$ is the sample mean of $\beta_i^{d_k}$'s, which is $\overline{\beta}_i^d = (1/K)\sum_{k=1}^{K} \beta_i^{d_k}$, and $\text{se}(\overline{\beta}_i^d)$ is the standard error of $\overline{\beta}_i^d$. Under the null hypothesis, the statistic Λ_i defined in (2.9) follows approximately standard normal distribution when k is large. P value for rejecting the null hypothesis at a significance level $\alpha = 0.05$ can be utilized to identify significant SNPs.

2.5. Application of the Multinomial Probit Model with SVD

2.5.1. Simulated Multinomial Ordinal Data

The validity of the proposed method was evaluated using simulated data sets. The procedure of data generation was composed of three steps: generating genotype data with certain genetic model, generating the latent variable, and defining the disease status variable by applying the predefined bin boundaries. The brief scheme of each step is as follows: we first generated 10 sets of the simulated genotype data under an additive genetic model, each set consists of 100 samples and 1000 SNPs. From (2.1), we can notice that the latent variable (z_i) consists of two parts: the expected value ($x_i\beta$) and the random error (ϵ_i). In order to generate the expected value, we assumed that, for each sample, 9 out of the 1000 SNPs (every 101th SNP, except the last one) contribute to disease status $x_i\beta = \beta_1 \cdot \text{SNP}_{101} + \beta_2 \cdot \text{SNP}_{201} + \cdots + \beta_1 \cdot \text{SNP}_{801} + \beta_2 \cdot \text{SNP}_{901}$, where β_1 and β_2 are set as -1 and 1, respectively. Hence, the latent variable can be obtained from the sum of the expected values ($x_i\beta$) and the random error generated from standard normal distribution. We then generated disease status variable (y_i) assuming 3 disease development stages. Therefore, when applying the proposed method to the simulated data sets, we would expect 9 strong signals corresponding to each of the 9 disease-associated SNPs. We also compared results obtained from the proposed method with that from single SNP analysis with multinomial ordinal probit model.



(a) Estimates of SNP effects by the proposed method (b) P values of testing SNP effects ($-\log_{10}(P \text{ value})$)

Figure 1: Analysis of the simulated data sets with multinomial ordinal probit model with SVD.

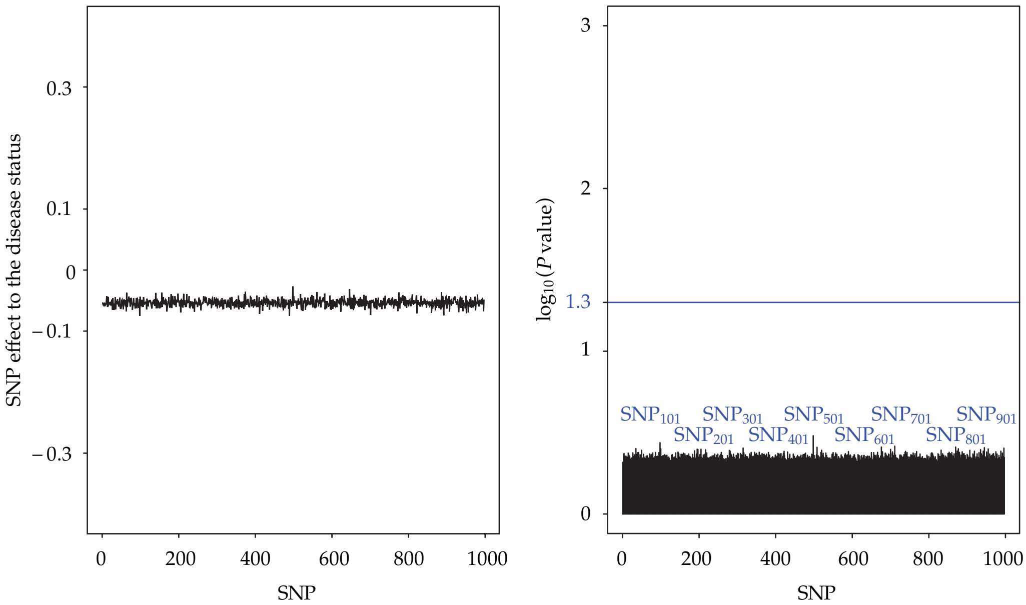
2.6. Mexican-American Coronary Artery Disease (MACAD) Study

We also applied the proposed method to study sample recruited through the Mexican-American Coronary Artery Disease (MACAD) study [8, 9]. The study population consists of probands who are Mexican American aged between 45 and 75 with coronary artery disease: spouses of probands, adult offspring (≥18), and their spouses. For the offspring generation, we performed oral glucose tolerance test and genotyped 132 SNPs in 32 genes selected based on a prior relationship to insulin physiology. The goal of the study herein was to identify genes involved in the development of IGT and/or IFG, where IGT was defined as a 2 hr glucose level between 140 and 199 mg/dL and IFG defined as a fasting glucose level between 100 and 125 mg/dL. In order to identify and compare genes affecting the development of IGT and/or IFG, we generated two study samples, for which each sample has 3 disease stages (D1) both 2 hr and fasting glucoses normal $(N/N)(n_1 = 60)$, IGT only (IGT/N) ($n_2 = 31$) and IGT and IFG (IGT/IFG) ($n_3 = 15$) (D2) both 2 hr and fasting glucoses normal $(N/N)(n_1 = 60)$, IFG only (N/IFG) ($n'_2 = 34$) and IGT and IFG (IGT/IFG) ($n_3 = 15$).

3. Results and Discussion

3.1. Simulated Multinomial Data

Figure 1 summarizes the results of association analyses when applying the multinomial ordinal probit model with SVD to the simulated data sets. All numbers shown in the figures are the average of the estimates obtained from the 10 simulated data sets. As mentioned previously, we expected 9 strong signals corresponding to the 9 SNPs designed to be associated



(a) Estimates of SNP effects by single SNP analysis

(b) P values of testing SNP effects ($-\log_{10}(P \text{ value})$)

Figure 2: Analysis of the simulated data sets with single SNP analysis.

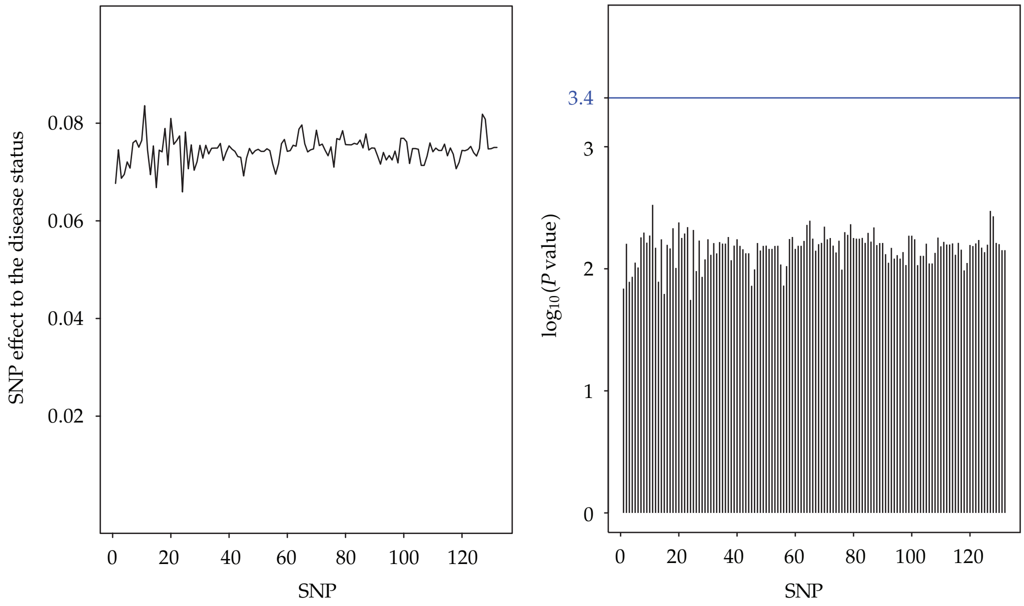
with disease development when generating the simulated data sets, and 9 were observed in our analysis. Similar results from the single SNP analysis were shown in Figure 2.

Figure 1(a) summarizes MLEs of SNP effects calculated with the multinomial ordinal probit model with SVD. The figure shows that almost all MLEs except 9 were between −0.1 and 0.1, while there were 9 large MLEs (4 around 0.3, 5 around −0.3) corresponding to the 9 SNPs contributed to disease status. Figure 1(b) gives P values in $-\log_{10}$ scale for testing SNP effects. The line in Figure 1(b) corresponds to significance level $\alpha = 0.05$. 9 SNPs were clearly separated from the rest and had $-\log_{10}(P \text{ value}) > 1.3$.

Figure 2(a) summarizes MLEs of SNP effects obtained by the single SNP analysis and shows that no signal was strong enough to be distinguished from all other signals. The P values are given in Figure 2(b) in $-\log_{10}$ scale. SNP_{501} in the middle of the figure had a relatively strong signal compared to all others. However, the $-\log_{10}(P \text{ value})$ was much less than 1.3, which corresponds to significance level $\alpha = 0.05$. Thus, no SNPs were identified as statistically significant from the single SNP analysis method. In contrast to the fact that no SNP was identified as statistically significant by the single SNP analysis, the multinomial ordinal probit model with SVD method was able to identify all 9 SNPs contributing to disease status as statistically significant at significance level $\alpha = 0.05$. These results indicated that the proposed method should be reliable for the analysis of large-scale genome-wide association data that have polytomous ordinal responses when $m \gg n$.

3.2. Mexican-American Coronary Artery Disease (MACAD) Study

We analyzed the data sets D1 and D2 (see methods) generated from a subsample of subjects recruited through a coronary artery disease proband in the Mexican-American Coronary



(a) Estimates of parameters by single-SNP analysis

(b) P values of testing SNP effects ($-\log_{10}(P \text{ value})$)

Figure 3: Analysis of genes for IGT/IFG through IGT pathway (Data Set D1) with single-SNP analysis.

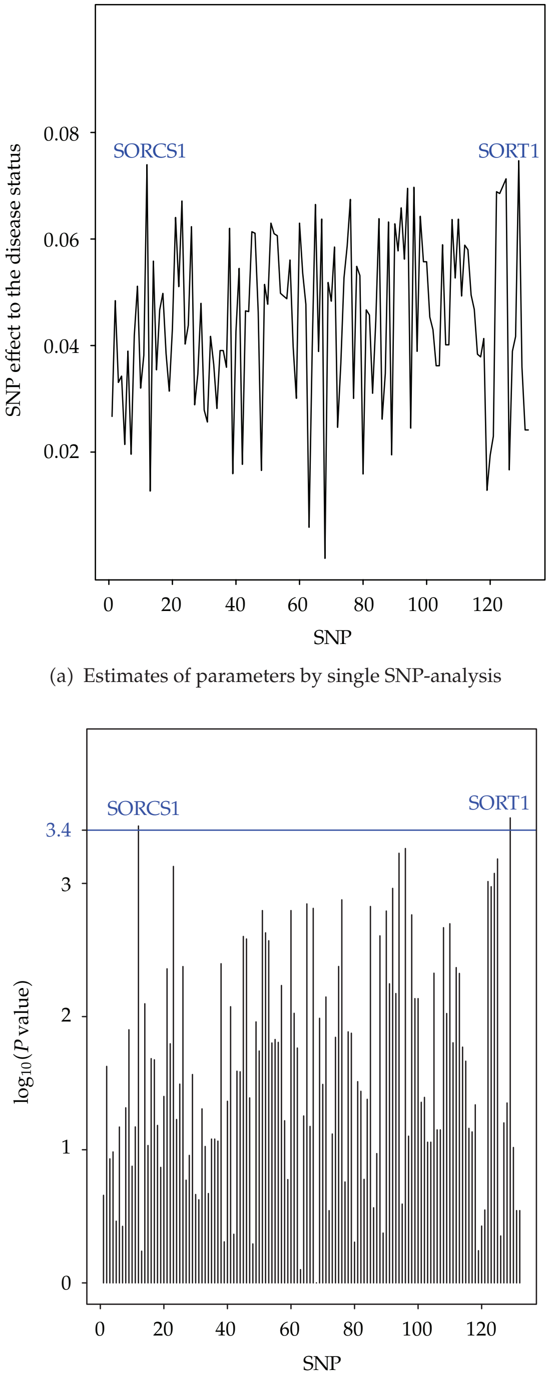
Artery Disease Project as described in the method section, using both the multinomial ordinal probit model with SVD method and the single SNP analysis method.

Figure 3 summarizes the analysis results with the data D1 (N/N-IGT/N-IGT/IFG) using the single SNP analysis. Figure 3(a) gives MLEs of SNP effects. Figure 3(b) plots P values of association analysis in $-\log_{10}$ scale. With Sidak correction, which is often used to correct multiple testing problem, the adjusted significance level should be $1 - (1 - \alpha)^{1/m}$, where α is significance level, and m represents the number of tests. Thus, the corrected $-\log_{10}(P \text{ value})$ threshold for significance level $\alpha = 0.05$ is 3.4, which corresponds to the line in Figure 3(b). We applied the adjusted significance level to the P values in Figure 3(b) since the P values are before correcting multiple testing problem. No SNP was identified as statistically significant (Figure 3(b)).

The data set D2 (N/N-N/IFG-IGT/IFG) was analyzed with the same method, and analysis results are given in Figure 4. Figure 4(a) plots MLEs of the SNP effects. Since P values in Figure 4(b) are before the the multiple testing correction, we used 3.4 as the $-\log_{10}(P \text{ value})$ threshold corresponding to 0.05 significance level after the multiple testing correction. Two SNPs corresponding to SORC1 and SORT1 were found significant.

We then also analyzed D1 and D2 with the multinomial ordinal probit model with SVD method. Figure 5 summarizes the analysis results for data D1. Figure 5(a) plots MLEs of SNP effects. Figure 5(b) plots P values in $-\log_{10}$ scale for testing SNP effects. The multiple testing correction does not need to be applied now since the method tests all SNPs simultaneously. With the 1.3 P value threshold, which corresponds to 0.05 significance level, we identified that 8 out of the 32 candidate genes (SORCS1, AMPD1, PPARα, AMPD2, PRKAA2, C5, TCF7L2, and ITR) were associated with the disease path defined in D1.

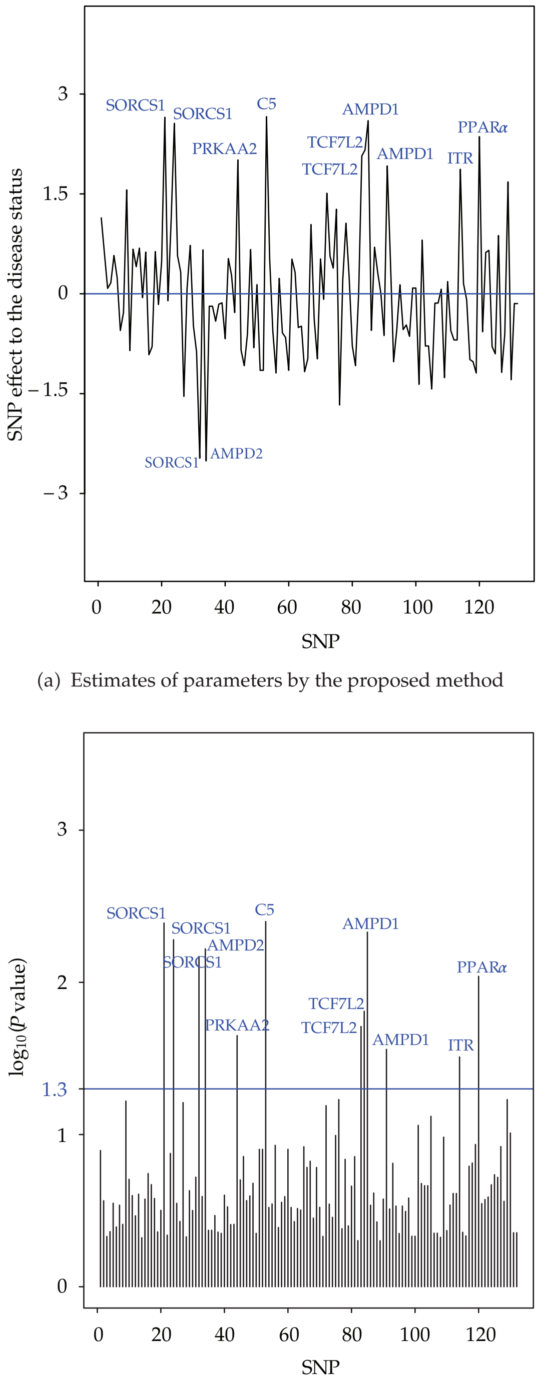
The multinomial ordinal probit model with SVD method was applied to data set D2 as well. The results are shown in Figure 6. In Figure 6(a), MLEs of the SNP effects were



(a) Estimates of parameters by single SNP-analysis

(b) P values of testing SNP effects ($-\log_{10}(P\text{ value})$)

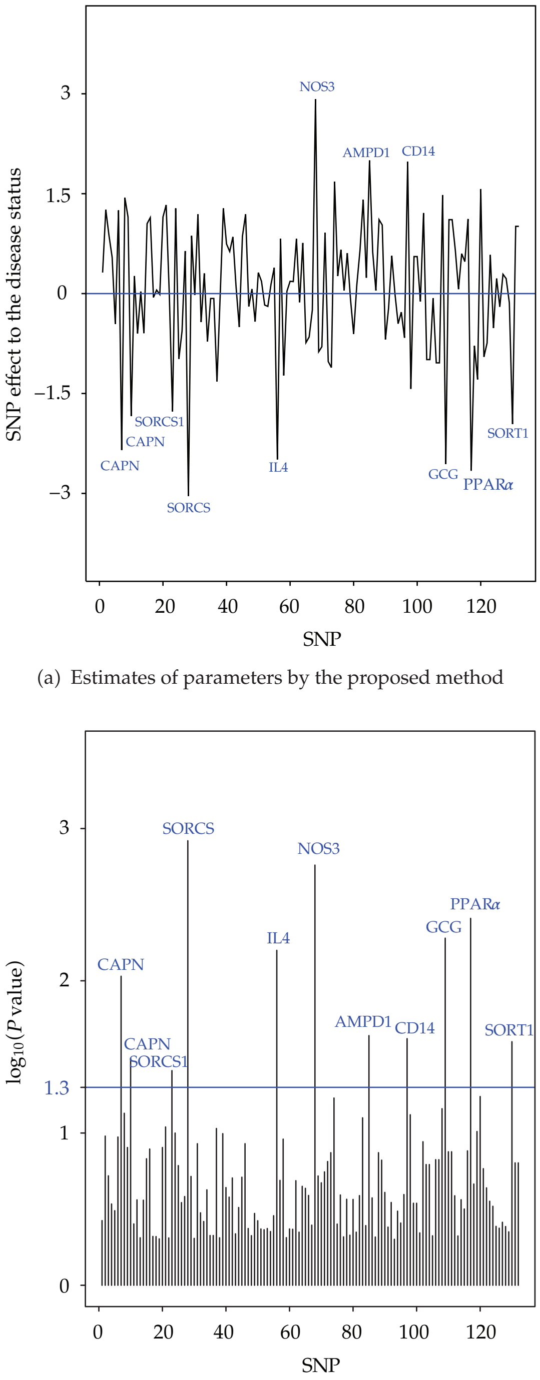
Figure 4: Analysis of genes for IGT/IFG through IFG pathway (Data Set D2) with single SNP-analysis.



(a) Estimates of parameters by the proposed method

(b) P values of testing SNP effects ($-\log_{10}(P\text{ value})$)

Figure 5: Analysis of Genes for IGT/IFG through IGT pathway (Data Set D1) with multinomial ordinal probit model with SVD.



(a) Estimates of parameters by the proposed method

(b) P values of testing SNP effects ($-\log_{10}$(P value))

Figure 6: Analysis of genes for IGT/IFG through IFG pathway (Data Set D2) with multinomial ordinal probit model with SVD.

summarized. Figure 6(b) plots the P values in $-\log_{10}$ scale for testing the SNP effects. It showed that 11 SNPs corresponding to 9 out of 32 candidate genes (SORCS1, AMPD1, PPARα, CAPN10, IL4, NOS3, CD14, GCG, and SORT1) have $-\log_{10}(P$ value) greater than the 1.3 P value threshold. From the analyses of D1 and D2, we found that SNPs in 3 genes (SORCS1, AMPD, and PPARα) were associated with both IGT and IFG; SNPs in 5 genes (AMPD2, PRKAA2, C5, TCF7L2, and ITR) were associated with IGT only; SNPs in 6 genes (CAPN, IL4, NOS3, CD14, GCG, and SORT1) were associated with IFG only. These results suggest that IGT and IFG may indicate different pathways to diabetes, with different genetic determinants.

Thus, using both simulated data and a real study sample, we demonstrated that multinomial ordinal probit model with SVD method can be utilized to identify associated markers involved in disease development when multidisease stages are considered. For relatively small size of data set used in the paper, which is 100 samples and 1000 SNPs for the simulation study, the computation took about less than 10 minutes to complete. However, the computation time might be a concern when applying this method to large data set, such as GAWS with millions of SNPs and thousands of samples.

Acknowledgments

This research was partly supported by Mexican American Coronary Artery Disease (MACAD) study Grant NIH-NHLBI HL 088457, Diabetes Endocrinology Research Center (DERC) Grant NIH-NIDDK DK063491, and UCLA Clinical and Translational Science Institute (CTSI) Grant NIH-NCATS UL1TR000124.

References

[1] S. Kwon, D. Wang, and X. Guo, "Application of an iterative Bayesian variable selection method in a genome-wide association study of rheumatoid arthritis," *BMC Proceedings*, vol. 1, supplement 1, article S109, 2007.

[2] E. I. George and R. E. McCulloch, "Approaches for bayesian variable selection," *Statistica Sinica*, vol. 7, no. 2, pp. 339–373, 1997.

[3] S. Kwon, J. Cui, K. D. Taylor, R. Azziz, M. O. Goodarzi, and X. Guo, "Application of Bayesian classification with singular value decomposition method in Genome-wide association study of rheumatoid arthritis," *BMC Proceedings*, vol. 3, supplement 7, article S9, 2009.

[4] W. H. Greene, *Econometric Analysis*, Prentice-Hall, Upper Saddle River, NJ, USA, 3rd edition, 1997.

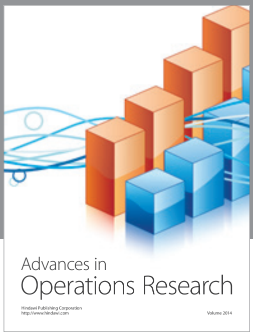
[5] J. Albert and S. Chib, "Bayesian analysis of binary and polychotomous response data," *Journal of American Statistical Association*, vol. 88, no. 422, pp. 669–679, 1993.

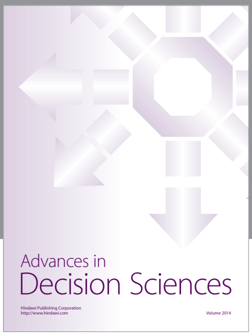
[6] J. Jansen, "Fitting regression models to ordinal data," *Biometrical Journal*, vol. 33, no. 7, pp. 807–815, 1991.

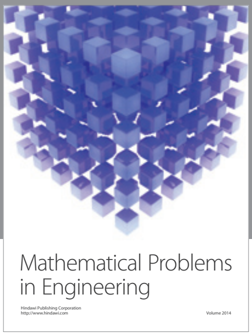
[7] F. A. Graybill, *Theory and Application of the Linear Model*, Duxbury Press, Belmont, Calif, USA, 1976.

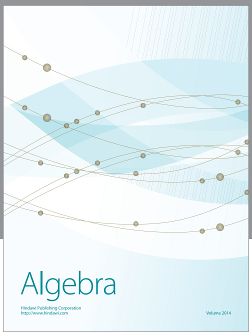
[8] M. O. Goodarzi, X. Guo, K. D. Taylor et al., "Determination and use of haplotypes: ethnic comparison and association of the lipoprotein lipase gene and coronary arter disease in Mexican-Americans," *Genetics in Medicine*, vol. 5, no. 4, pp. 322–327, 2003.

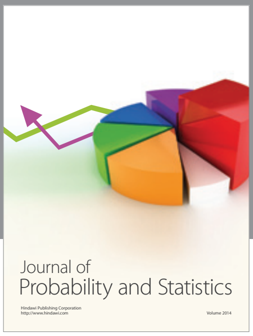
[9] M. O. Goodarzi, X. Guo, K. D. Taylor et al., "Lipoprotein lipase is a gene for insulin resistance in Mexican Americans," *Diabetes*, vol. 53, no. 1, pp. 214–220, 2004.


Advances in
Operations Research

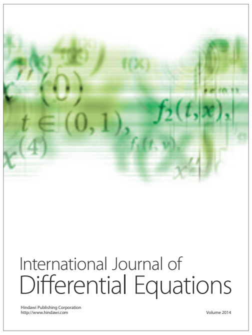

Advances in
Decision Sciences

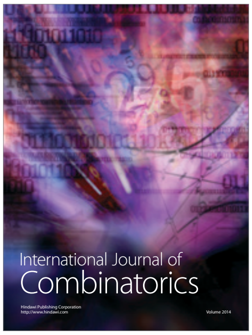

Mathematical Problems
in Engineering

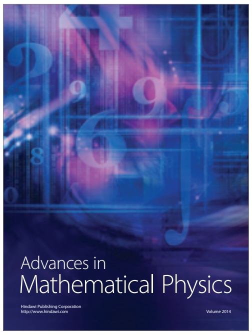

Algebra

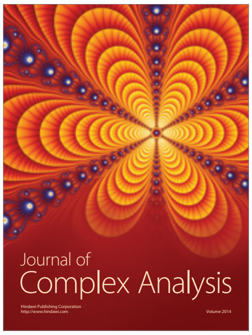

Journal of
Probability and Statistics

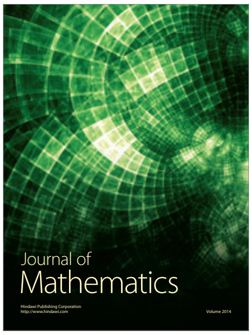
The Scientific
World Journal

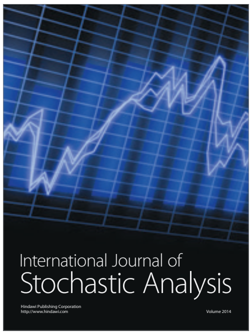

International Journal of
Differential Equations

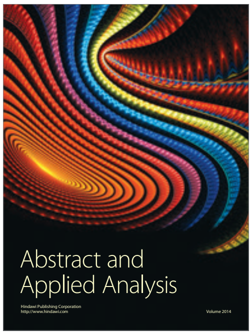

International Journal of
Combinatorics

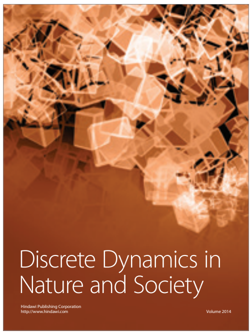

Advances in
Mathematical Physics

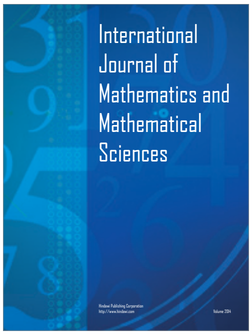

Journal of
Complex Analysis


Journal of
Mathematics


International Journal of
Stochastic Analysis

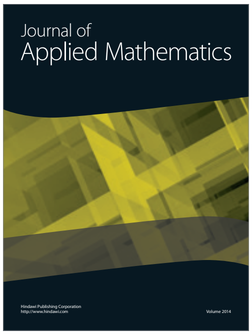

Abstract and
Applied Analysis


Discrete Dynamics in
Nature and Society


International
Journal of
Mathematics and
Mathematical
Sciences

Journal of
Discrete Mathematics

Journal of
Function Spaces


Journal of
Applied Mathematics

Journal of
Optimization