

ゲノムワイド関連解析データを用いた複雑疾患の階層混合モデル解析

西野穰^{1,4}, 高地雄太^{2,4}, 野間久史^{3,4}, 重水大智^{2,4}, 森園隆^{2,4}, 角田達彦^{2,4}, 松井茂之^{1,4}

¹名古屋大学大学院, ²理化学研究所, ³統計数理研究所, ⁴科学技術振興機構 CREST

リウマチや多くの精神疾患などの複雑疾患の発症には、多くの遺伝因子が寄与していると考えられている。これまで、ゲノム全域に渡る数十万以上の一塩基多型 (SNP) の遺伝子型と疾患形質との関連を調べるゲノムワイド関連解析 (GWAS) によって多くの疾患関連 SNP が同定されてきた。しかし、多くの疾患では、ゲノムワイドに有意な SNP を集めても、家系分析から推定される遺伝率 (heritability=遺伝的寄与度) のごく一部しか説明できないこと (missing heritability) が指摘されている[1]。一方、(GWAS で遺伝子型決定に用いる) SNP チップ上の全 SNP を同時に用いて推定される遺伝的寄与率 (SNP heritability) は、遺伝率の相当な割合を占める場合が多い。これはチップ上には小さな効果を持つ未知の SNP が数多く存在する事を示唆する[2]。本研究は、階層混合モデル解析[3]に基づいて、このような小さな効果をもつ SNP の評価を可能とする解析法の確立を目指す。

目的変数としてケース (患者) を $z = 1$ 、コントロール (正常) を $z = 0$ 、説明変数として各 SNP の遺伝子型 AA: $x = 0$, Aa: $x = 1$, aa: $x = 2$ としたロジスティック回帰を m 個の SNP 毎に行い、 j 番目の SNP についての対数オッズ比の最尤推定値 $Y_j = \hat{\beta}_j$ とその分散 $\hat{V}_{\hat{\beta}_j}$ を得たとする。いま、 j 番目の SNP の確率分布に対して次の階層混合モデルを仮定する：

$$f_j(y_j) = \pi f_{0j}(y_j) + (1 - \pi) f_{1j}(y_j)$$

ここで、 f_{0j} は null (関連なし) のコンポーネントに対応し $y_j \sim N(0, \hat{V}_{\hat{\beta}_j})$ と指定する。 f_{1j} は、non-null (関連あり) のコンポーネントに対応し $y_j | \beta_j \sim N(\beta_j, \hat{V}_{\hat{\beta}_j})$, $\beta_j \sim g_1$ という階層構造を指定する。効果サイズの分布 g_1 は指定しない。 π と g_1 の推定は、EM アルゴリズムにより行う (経験ベイズ推定) [3]。これより、小さい効果も含め各 SNP の効果をベイズ的に推定でき、効果サイズの全体像を捉える事が可能となる。

モンテカルロシミュレーションによって各 SNP の効果サイズの事後平均は、効果サイズの全領域に渡りほぼ不偏である事が示された。また、遺伝的寄与率の推定や発症の予測で用いる polygenic score (効果サイズの推定値で重み付けしたリスクアレル数の和) において、重みに事後平均を用いると精度が改善することが示された。本報告では、上記の内容に加え、リウマチの公開データを用い、実際の問題に対する本手法の有効性を示す。

[参考文献]

1. Manolio, T.A., et al. Finding the missing heritability of complex diseases. Nature 2009; 461: 747-753.
2. Lee, S. H., et al. Estimating missing heritability for disease from genome-wide association studies. The American Journal of Human Genetics 2011; 88: 294-305.
3. Matsui, S. and Noma, H. Estimating effect sizes of differentially expressed genes for power and sample - size assessments in microarray experiments. Biometrics 2011; 67: 1225-1235.