

集団構造化補正、Genomic Control法のフィッシャー正確確率検定への応用

The Application of Genomic Control Method, the Correction of Population Structure, to Fisher's Exact Test.

岡田随象・山田亮

東京大学医科学研究所 ヒトゲノム解析センター ゲノム機能解析分野

【目的】ゲノムワイド関連解析においては、対象集団の構造化が統計量のインフレーションを引き起こす事が指摘されている。Genomic Control (GC) 法は補正手段の一つであり、帰無仮説を満足すると想定される多数の χ^2 統計量の分布を理論的 χ^2 分布の定数倍として近似する手法である。期待度数が低い分割表においては χ^2 検定に代えてフィッシャー正確確率検定が適用されるが、後者に対するGC法の応用は示されていない。フィッシャー正確確率検定統計量は帰無仮説下で保守的な分布をとるため、応用に際しては帰無仮説下での統計量分布を推定する手法が必要と考えられた。

【方法】我々は帰無仮説下におけるフィッシャー正確確率検定統計量の分布を推定するため、ジェノタイプデータより得られた2x2分割表の周辺度数に基づき、検定統計量の理論的な分布を推定する手法を採用した。

【結果】帰無仮説下におけるフィッシャー正確確率検定統計量の分布に対する良好な推定結果が得られた。

序文

対象集団内に存在する集団構造化はケース・コントロール間に多型頻度の差を生じるため(Fig.1)、観測統計量の全体的な偏移が引き起こされる。GC 法は χ^2 検定による χ^2 値の中央値と $p=0.5$ となる χ^2 値 (= 0.455, df=1) の比を用いて χ^2 値の偏移を補正する手法である。^[1-2]

期待度数の少ない分割表においては χ^2 検定は不正確なP値を与えるため、フィッシャー正確確率検定が推奨される。フィッシャー正確確率検定P値は保守的な値をとり、帰無仮説下における中央値は $p=0.5$ より大きい値となる^[3] (Fig.2)。我々は同検定に対するGC法を、観測検定量の分布を集団構造化が無い状態における分布に近似する手法と定義した。

方法

分割表における内部度数の分布は、帰無仮説下において周辺度数に基づく超幾何分布に従う。ジェノタイプデータより生成されたallele型の各2x2分割表に対して、固定された周辺度数に対する内部度数の確率分布を、各内部度数から得られるフィッシャー正確確率検定P値と対応付ける事により、帰無仮説下での統計量分布を推定した。(Fig.3)

集団構造化有り・無し二種類のジェノタイプデータを作成した(サンプル数(N)=200, SNP数=100,000, マイナーアレル頻度(MAF)=0.01-0.5)。両データより得られたallele型2x2分割表からフィッシャー正確確率検定P値を算出した。集団構造化が存在するデータに対して、帰無仮説下におけるフィッシャー正確確率検定P値の分布を推定し、構造化が存在しないデータにおける分布と比較した。

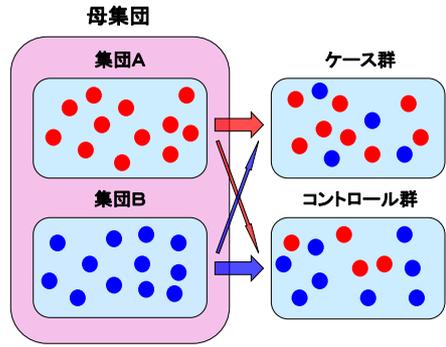


Fig. 1: 集団構造化によるケース・コントロール群間の乖離

母集団中の異なる集団(A・B)に属するサンプルが異なる比率でケース・コントロール群に分配された場合、**集団構造化に基づく乖離が群間に生じる。**

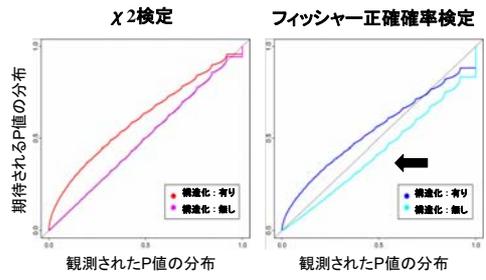


Fig. 2: χ^2 検定・フィッシャー正確確率検定P値のQQ-plot.

構造化有り・無し二種類のデータにおける両検定P値のQuantile-Quantile plot。

・共に集団構造化の存在下でP値の偏移を生じている。

・フィッシャー正確確率検定は χ^2 検定と比較して保守的である(黒矢印部)。

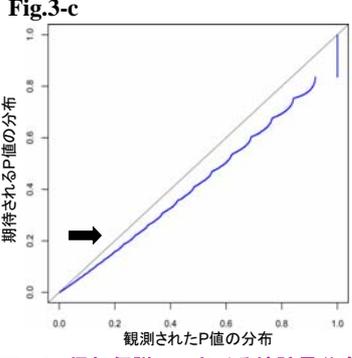
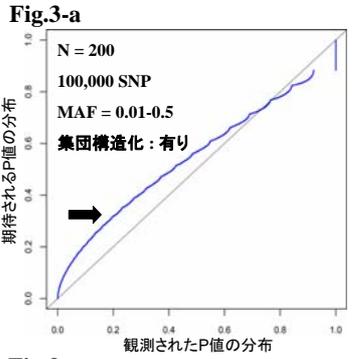
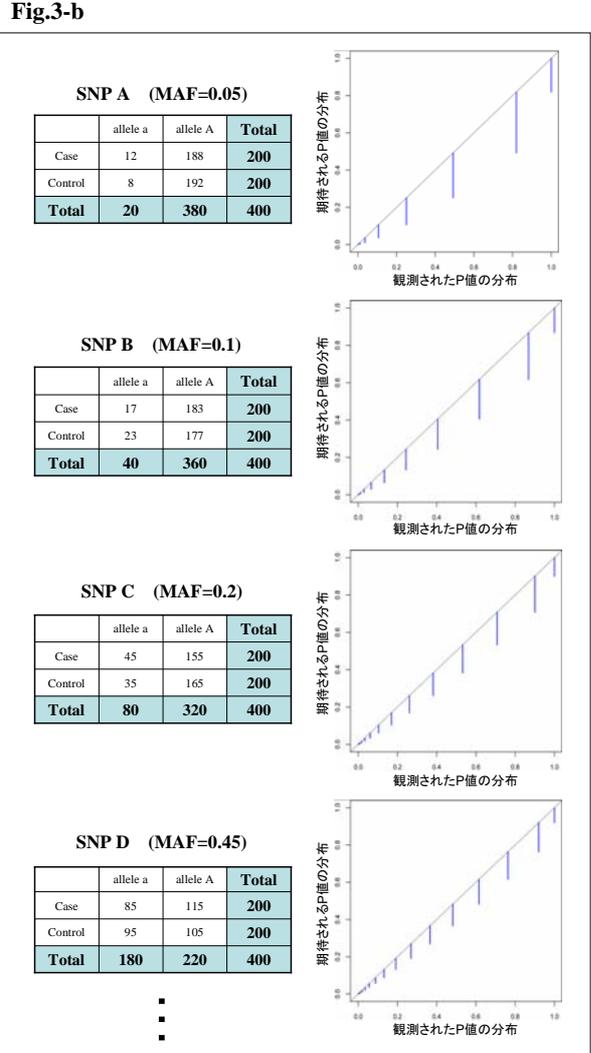


Fig. 3: 帰無仮説下における統計量分布の推定

集団構造化を含むデータの、帰無仮説下におけるフィッシャー正確確率検定P値の分布の推定。

- ・集団構造化により元データのP値の分布は偏移している(Fig.3-a、矢印部)。
- ・各SNPに対し、allele型2x2分割表の周辺度数に基づく帰無仮説下の分布を推定する(Fig.3-b)。
- ・全SNPに対し推定された分布を足し合わせる事により、元データの帰無仮説下におけるP値の分布を推定した。

⇒P値の偏移が消失した(Fig.3-c、矢印部)。



結果

- ・集団構造化が存在するデータより推定された帰無仮説下の分布において、P値の偏移の消失が認められた。
- ・推定された帰無仮説下の分布と、構造化の存在しないデータから計算された分布との間に、±2SDの範囲において有意な差を認めなかった。

考察

- ・フィッシャー正確確率検定P値の帰無仮説下における分布に対する良好な推定結果が得られた。
- ・推定された帰無仮説下の統計量分布を採用することにより、GC法のフィッシャー正確確率検定への応用が可能であると考えられた。

引用文献

- (1) Delvin, B. & Roeder, K. Genomic control for association studies. *Biometrics* 55, 997-1004 (1999).
- (2) Delvin, B. & Roeder, K. Wasserman L. Genomic control, a new approach to genetic-based association studies. *Theor. Popul. Biol.* 60, 155-66 (2001).
- (3) Graham, J.G. Fisher's Exact Test. *J. R. Statist. Soc. A.* 155, 395-402 (1992).

アプリケーション

<http://func-gen.hgc.jp/member/yo/FisherGC.tar>
よりダウンロード可能。
ご質問は yokada-tyk@umin.ac.jp まで。