

2M-04 シミュレーションによる Hardy-Weinberg 平衡の検定方式の評価

野崎康行[†] 永井淳[‡] 田村卓郎[†] 中重亮[†]

[†] 日立ソフトウェアエンジニアリング (株) [‡] 岐阜大学医学部法医学教室

1. はじめに

DNA 多型の研究のひとつとして、DNA 鑑定で用いられる DNA 観察集合の遺伝的安定性の検証について検討を行った。DNA 観察集合とは、DNA 鑑定の際に用いる無作為に母集団から抽出したサンプルデータのことであり、信頼できる鑑定結果を示すためには、DNA 観察集合に蓄積されたデータが、母集団を反映したものである必要がある。そこで、DNA 観察集合が母集団を反映することを示すための検定方式について、既存方式の有効性の評価を行った。

2. DNA 鑑定の方法

DNA 配列中には、部分的に STR とよばれる 2~5 塩基の同じ配列が連続して繰り返し現れる領域が存在する。そしていくつかの STR 領域では、個体間で反復回数にばらつきがあることが知られている。

DNA 鑑定を行うには、その準備段階として、STR 領域での部分配列の反復回数の頻度分布を調べるために、あらかじめサンプルで反復回数を計測しておかなければならない。この計測値を蓄えたものを、本報告では DNA 観察集合とよぶ。ヒトは 2 倍体なので、1 サンプルにつき、二つの反復回数が計測される。

鑑定では、特定個人のサンプルから STR 領域での反復回数を計測し、DNA 観察集合から該当するサンプルが得られる確率を求める。STR 領域間の確率は独立とみなしてよいことが知られているので、複数の STR 領域での反復回数を観測すれば、それぞれの STR 領域での確率の積をとることによって、該当するサンプルが得られる確率をより小さくすることが可能になる。このように DNA 鑑定とは、確率という客観的な尺度を持って、個人の同定を示す手法である。

以後、反復回数自体は問題とせず、STR 領域での配列

を単に塩基配列の差に基づく DNA 多型とし、この配列のことをアリルとよぶ。またサンプルがもつ一対のアリルの組合せのことを遺伝子型とよぶ。

3. DNA 観察集合の正当性の検証

鑑定を正しく行うには、DNA 観察集合でのアリルの確率分布が正確で安定していなければならない。つまり、DNA 観察集合が母集団をうまく反映したものであるかを検定する必要がある。

DNA 観察集合が母集団を反映しているかどうかは、Hardy-Weinberg 平衡という特性に着目した検定で行う。サンプル数を n 、ある STR 領域でのアリルの総数を m 、アリルを A_1, A_2, \dots, A_m 、遺伝子型 $A_i A_j$ の DNA 観察集合での観測値を f_{ij} 、アリル A_i の現れる度数を f_i とおくと、Hardy-Weinberg 平衡とは以下の式を満たすことである。

$$\begin{cases} f_{ii} = f_i^2 / 4n \\ f_{ij} = f_i f_j / 2n \quad (i < j) \end{cases} \quad \text{----- (1)}$$

式 (1) により f_i, f_j から算出される $f_{ij} (i \leq j)$ を期待値として、 $\tilde{f}_{ij} (i \leq j)$ と表す。

DNA 観察集合が Hardy-Weinberg 平衡を満たしているかどうかは、実際の観測値 f_{ij} と期待値 \tilde{f}_{ij} を用い、自由度 $\frac{m(m-1)}{2}$ の χ^2 検定または尤度比検定で検定する。この他にも Homozygosity test や Exact test を用いる方法もある。法医現場における DNA 観察集合の検定では、通常有意水準を 5% に設定するので、本報告もこれに従った。

χ^2 検定や尤度比検定は、観測値と期待値のばらつき度合いを表す χ^2 値の頻度分布が、 χ^2 分布に近似されることを根拠とした検定手法である。しかし遺伝子型の種類に比べサンプルの数が少ない場合は χ^2 分布に適合しにくい。そこで「5 の法則」を適用して、少ない遺伝子型になるよう DNA 観察集合を再構成した後に検定を行えばよいことが経験的に知られている。

5 の法則とは、DNA 観察集合での期待値が次の二つの条件を満たすと、遺伝子型の種類に比べてサンプルの数が少ない場合でも χ^2 分布に適合しやすくなるというもの

Evaluation of Tests for Hardy-Weinberg Equilibrium by Simulation

Yasuyuki Nozaki[†], Atsushi Nagai[‡], Takuro Tamura[†] and Ryo Nakashige[†]

[†] Hitachi Software Engineering Co., Ltd.

[‡] Department of Legal Medicine, Gifu University School of Medicine

である。

- (1) どの遺伝子型でも期待値が1以下にならない。
- (2) 全体の2割以上の遺伝子型で期待値が5以下にならない。

5の法則を適用するとは、上の二つの条件を満足するように、観測数の少ないアリルを隣接するアリルと逐次合併してゆき、条件が成立するまでグループ化することである。

4. 検出力の検証

しかし、5の法則は経験則であるため以下のような問題が生じる。

- (1) 5の法則を適用したときにそれがどの程度 χ^2 分布に近似されるかが明確でない。
- (2) 5の法則を適用すべきサンプル数、遺伝子型の種類の数に対する条件が明確でない。

そこで、シミュレーションによって遺伝子型の度数 f_{ij} を大量に生成し、 χ^2 値の度数分布を作成し、それらが χ^2 分布に近似されるかどうかを定量的に調べた(図1)。

度数分布において χ^2 分布の5%棄却域に入る割合(実験確率)を求めると、度数分布と χ^2 分布との差の最大値を求めることの二点で評価を行った。

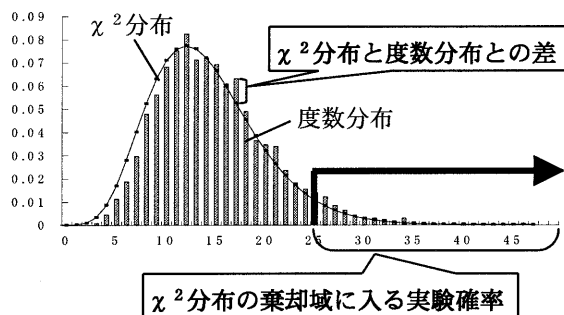


図1 シミュレーションによって作成した度数分布

5. シミュレーション結果

様々なサンプル数とアリル数に対して、実験確率と、 χ^2 分布との差の最大値を求め、これを1,000回実行することで、それぞれ平均値と標準偏差を算出した。これらを全データ(5の法則を考慮していない場合)と、5の法則が成立するものを選別した場合で区別して評価した。次の二つの場合に分けて考察した。

- (1) サンプル数が固定(=400)、アリル数が可変(=4~9)の場合
- (2) サンプル数が可変(=50~10,000)、アリル数が固定(=7)の場合

変化の推移を把握するために、可変部分を横軸にとり、平均値を折れ線グラフで表した。図2、図3は χ^2 検定の実験確率をサンプル数固定、アリル数可変(1)とサンプル数可変、アリル数固定(2)の条件の下で比較したものである。シミュレーションで、平均値がほぼ χ^2 分布の棄却域の値(0.05)をとっていることがわかる。

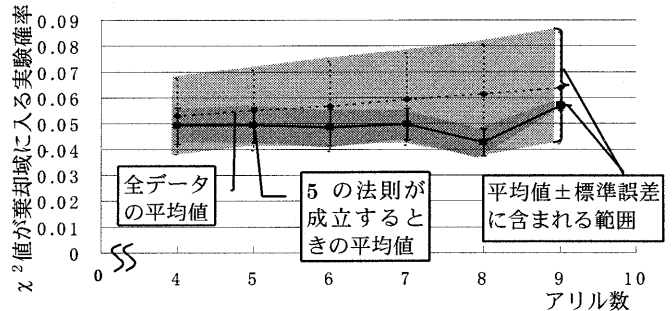


図2 アリル数可変での実験確率の推移

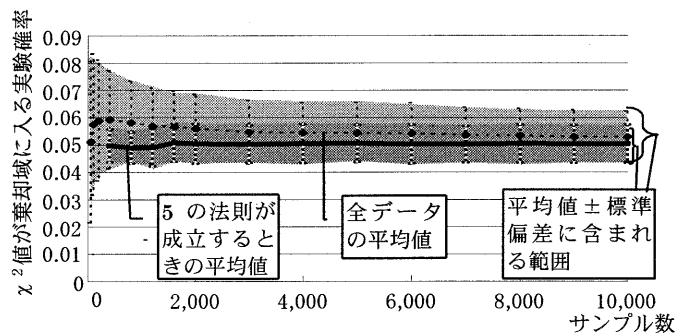


図3 サンプル数可変での実験確率の推移

6. まとめ

シミュレーションの結果から以下の結論を得た。

- (1) 5の法則を適用すると χ^2 分布によく近似される傾向がある。尤度比検定においては、サンプル数が小さいと近似されにくいので、5の法則適用後もさらにグループ化をすすめてアリル数を充分小さくする必要がある。
- (2) サンプル数が3,000以下のとき、5の法則を適用していないときは平均値の分布にばらつきが大きい。したがって3,000以下のときは5の法則を適用すべきである。

参考文献

- [1] Evett, I. E. and Weir, B. S. : Interpreting DNA evidence, Sinauer Associates, Inc. (1998)
- [2] 勝又義直, 水谷正樹, 野澤秀樹, 打樋利英子, 山本敏充 : 日本人集団におけるハーディ・ワインベルグ平衡の検定法について, 日本法医学雑誌(日本医誌), Vol.53, pp.218-226 (1999-6)
- [3] 永井淳, 山田定男, 中村功, 竹内康雄, 大谷勲 : 蛍光イメージアナライザーを用いた8種類のSTR座位の同時検出と日本人集団における各アレルの出現頻度, DNA多型, Vol.6, pp.179-183 (1998)