

DNA microarray データを検定しやすく変換する

Transformation of DNA microarray data for statistical tests

井上 真郷 (PY)^{1,4}, 西村 信一^{1,5}, 堀 玄², 甘利 俊一¹, 斉藤 美知子³, 吉原 良浩³, 中原 裕之¹

Masato Inoue^{1,4}, Shin-ichi Nishimura^{1,5}, Gen Hori², Shun-ichi Amari¹,

Michiko Saito³, Yoshihiro Yoshihara³, Hiroyuki Nakahara¹

理化学研究所 脳科学総合研究センター

{¹脳数理研究チーム, ²脳信号処理研究チーム, ³シナプス分子機構研究チーム},

⁴京都大学大学院医学研究科 耳鼻咽喉科・頭頸部外科, ⁵東京大学大学院医学系研究科 耳鼻咽喉科学専攻

{¹Lab. for Mathematical Neuroscience, ²Lab. for Advanced Brain Signal Processing,

³Lab. for Neurobiology of Synapse}, RIKEN Brain Science Institute,

⁴Dept. of Otolaryngology, Head and Neck Surgery, Graduate School of Medicine, Kyoto Univ.,

⁵Dept. of Otorhinolaryngology, and Head and Neck Surgery, Graduate School of Medicine, Univ. of Tokyo

E-mail: minoue@brain.riken.go.jp

DNA microarray は、任意の生体組織について、数千～数万個の予め選択した遺伝子がどの程度発現しているかを同時に調べる事ができる画期的な測定システムであり、これにより遺伝子研究が活性化する一方、大量のデータから統計的により正確な推測をする必要性が高まっている。

典型的な課題は、複数条件下で測定を繰り返し、条件間で発現量に差がある遺伝子を有意度順に並べるといったものである。今、1 万個の遺伝子について、各条件下で数回測定を繰り返したデータがあるとすると、条件間で発現の変化があるかどうかは分散分析を行えばよいし、薬物量との関係を調べるには回帰分析が使える。

しかし、これらの検定や解析は一般に「正規性」と「等分散性」を前提とする。即ち、各条件下での測定データは正規分布に従い、どの条件下でも共通の母分散を持っていなければならない。これが各遺伝子についてそれぞれ成立する事が求められる。分散分析はこれらの前提が少々崩れていても頑健性があり、また、non-parametric 検定という選択肢もあるが、この二つの性質は、他の多くの解析手法でも陰に陽に前提としており、分散分析に限らず重要である。

一般に、生物学的データは対数をとると正規分布に近くなる事が多いため（対数正規分布）、各観測値に対して対数変換が行われる。これは DNA microarray データにもだいたい当てはまるが、1) 非正値が含まれる、2) 変動係数が平均が 0 付近で増大する、という癖があり、単純な対数変換では

1)非正値のため全てのデータを変換できず、2)正値のみを変換したとしても、（変動係数が一定でないため）変換後の分散が一定にならないという問題があった。

Oligonucleotide array (GeneChip®)と呼ばれるタイプの DNA microarray について、観測値を確率変数 X とすると、

$$\text{Var}[X] = f(E[X]) = a(E[X] - b)^2 + c$$

という関係が理論上成立する。与えられた発現量データの、各遺伝子の各条件での平均と分散を用いて共通のパラメータ a, b, c を決定する事ができ、この、平均から分散を推定する関数 f を用いると、分散を近似的に 1 に正規化する変換 g を得る事ができる。

$$g(X) = \frac{1}{\sqrt{a}} \ln \frac{1}{2} \left(\sqrt{(X - b)^2 + c/a} + X - b \right)$$

また、 g を求めるのに、「 X がある閾値以下の場合には別処理を行う」といった事前のパラメータ設定は一切必要としない。

公開データを用いて、本手法の検証を行ったところ、等分散性を改善した。また、正規性は変換前でも良好で、変換により悪化しなかった。従って、この変換は従来の対数変換に代わるもので、多くの解析の前処理として有用であると思われる。