

シンポジウム資料 (2002.3.28)

一般化線形モデル (Generalized Linear Models)

多くの統計的検定や推定は、これを統一的な枠組みの下に整理しようとする試みはいくつかあったが、その中で次第に広く使われてきているのが一般化線形モデルである。

一般化線形モデルは次の3つの構成要素からなる

- ・ 連結関数 (link function)
- ・ 誤差構造 (error structure) またはランダム成分 (random component)
- ・ 線形予測子 (linear predictor)

独立変数の一次式が線形予測子、線形予測子とこれに対応する従属(目的)変数の期待値の関係を式にしたものが連結関数、従属変数が期待値のまわりにどうばらつくかが誤差構造である。

さまざまな方法は、連結関数と誤差構造の組み合わせがちがうものと考えると整理して理解でき統一的に扱える。

例：ロジスティック回帰

ロジスティック回帰では、線形予測子と $\log\{p/(1-p)\}$ の期待値が等しい。また、誤差は二項分布である。

そこで、一般化線形モデルでは、ロジスティック回帰を連結関数がlogit link、誤差の構造が二項分布と表現することになる。

例：線形(一次式)の重回帰

重回帰の場合、従属変数の期待値 = 独立変数の一次式すなわち線形予測子なので、連結関数はそのままということになる。これをidentity linkと呼ぶ。誤差の構造は、正規分布(等分散)である。

そこで、一般化線形モデルでは、重回帰を連結関数がidentity link、誤差の構造が正規分布と表現することになる。

また、分散分析のかなりの部分も同様にidentity linkで誤差の構造が正規分布である(ただし、独立変数の値が離散的)。

連結関数としてよく使われるものには、他に、log link (対数)、reciprocal link (逆数)、complimentary log-log link、exponent link (べき乗)、square root link (平方根) などがある。

誤差としては、負の二項分布 (ポアソン分布に対応するが、分散は平均値と等しくない)、ガンマ分布 (変動係数が一定)、ベータ二項分布 (二項分布に対応するが、それよりも分散が大きい) など使われる。

推定と検定：標準的に使われるのは、推定については最尤法、検定については尤度比検定である。ただし、誤差が正規分布で等分散であるときには、分散分析や回帰でよく使われる F 検定も用いられる。尤度比検定は小標本のときには注意が必要である。

参考になる文献

Generalized linear models. 2nd ed.
P. McCullagh and J.A. Nelder
Chapman and Hall. 1989年

GLIM for Ecologists.
M. J. Crawley
Blackwell Scientific Publications. 1993年

(粕谷)