

カウント型モデル用推定コマンド

ここで紹介するモデルにおいては従属変数が事象の発生回数を表すものであることが前提となります。代表的なモデルとしてはポアソン回帰モデルと負の二項回帰モデルがあります。

1. ポアソン回帰モデル (Poisson regression model) をフィットさせるコマンドは `poisson` ([R] `poisson` (*mwp-087*) 参照) であるわけですが、そのモデル式は

$$y_j \sim \text{Poisson}(\mu_j) \quad (9.1)$$

ただし $\mu_j = \exp(\mathbf{x}_j\boldsymbol{\beta} + \text{offset}_j)$

のように表現されます。すなわち事象の発生回数を表す従属変数の観測値 y_j は平均を μ_j とするポアソン分布に従い、かつ μ_j は共変量ベクトル \mathbf{x}_j によってパラメータ化されるというわけです。

2. これに対しポアソン分布よりも大きなバラツキを持ったモデル (overdispersion model) を想定したい場合には負の二項回帰モデル (negative-binomial regression model) が用いられます。このモデルをフィットさせるコマンドは `nbreg` ([R] `nbreg` 参照) で、そのモデル式は

$$y_j \sim \text{Poisson}(\mu_j^*) \quad (9.2)$$

ただし $\mu_j^* = \exp(\mathbf{x}_j\boldsymbol{\beta} + \text{offset}_j + \nu_j)$ $\nu_j \sim \text{Gamma}(1/\alpha, \alpha)$

のように記述されます。 α は overdispersion parameter と呼ばれ、その値が大きいほどバラツキは大きなものとなります。

3. Truncation (切断) を伴うカウント型モデルの場合、アウトカムを意味するカウント値はある閾値 (通常は 0) を越えた場合にのみ観測されることとなります。 `poisson`, `nbreg` に対応する形で `tpoisson` ([R] `tpoisson` 参照) `tnbreg` ([R] `tnbreg` 参照) というコマンドが用意されています。
4. Zero inflation を伴うカウント型モデルの場合、0 というカウント数が通常期待される値よりも大きな値となります。この超過分はプロビット、またはロジットプロセスによってモデル化されます。すなわち、このプロセスによって正のアウトカムがもたらされた場合には通常のカウント型プロセスが実行されますが、そうでない場合にはカウントは 0 とされます。該当するコマンドには `zip` ([R] `zip` 参照) `zinb` ([R] `zinb` 参照) の 2 種類があります。

■