応用システム工学
第7回　指数型分布族と一般化線形モデル

平成22年7月16日

統計的推測

統計的推測

- 信頼区間(区間推定)
  - 検定結果:信頼区間が値を含むかどうか
  - 検定精度:信頼区間の幅
- 仮説検定
  - 2つの関連するモデルのデータへの適合度の判定
    - 2つの一般化線形モデル（同一確率分布，連結関数）
      - パラメータの少ないモデル→帰無仮説
      - パラメータの多いモデル→対立仮説
      - 簡単なモデルが複雑なモデル程度に当てはまるかを評価する
  - 要約統計量(適合度統計量)を用いた当てはめの適合性
    - 尤度関数の最大値
    - 対数尤度関数の最大値
    - 平方和基準の最小値
    - 残差に基づく複合型の統計量

仮説検定

- 作業仮説・実験仮説を対立仮説H1
- 作業仮説を否定する仮説を帰無仮説H0
- 帰無仮説を仮説検定の対象とする。
  - 帰無仮説が棄却されると対立仮説が支持される。
  - 帰無仮説が支持される場合
    - 真に対立仮説が誤っている
    - 対立仮説は正しいが，標本の大きさが十分でなく，帰無仮説を積極的に棄却できない

仮説検定のプロセス

- モデルを決める
  - 帰無仮説H0に対応するモデルM0
  - 対立仮説H1に対応するモデルM1
- モデルに対する適合統計量を求める
  - M0→G0, M1→G1
- 適合のよさの比較・評価を行う
  - G1-G0またはG1/G0
- 対立仮説G1≠G0に対する帰無仮説G1=G0の検定
  - G1-G0の標本分布を評価
    - G1=G0が棄却されない→H0は棄却されない→M0がより良いモデル
    - G1=G0が棄却される→H0は棄却される→M1がより良いモデル

一般化線形モデルの標本分布で当てはまりを評価

スコア統計量の標本分布

- 一般化線形モデル
  - 互いに独立な確率変数Y1,・・・,YN
  - パラメータβ
  - 期待値 $\mu_i = E[Y_i]$
  - 連結関数 $g(\mu_i) = x_i^T \beta = \eta_i$
  - スコア統計量 $U_j = \frac{\partial l}{\partial \beta_j} = \sum_{i=1}^{N} \left[\frac{(Y_i - \mu_i)}{\mathrm{var}(Y_i)} x_{ij} \left(\frac{\partial \mu_i}{\partial \eta_j} \right) \right], j = 1, \cdots, p$
  - スコア統計量の期待値
    - $E[Y_i] = \mu_i$ より $E[U_j] = 0, j = 1, \cdots, p$

スコア統計量の標本分布

- 情報行列 $\mathfrak{I}$
  →スコア統計量の分散共分散行列
  – 要素

$$\mathfrak{I}_{jk} = E\left[U_j U_k\right]$$

- 単一のパラメータ β に対するスコア統計量の標本分布について考える

スコア統計量の標本分布

- 尤度関数 $l = \sum l_i$

$$l_i = \log f\left(Y_i, \beta\right)$$

- スコア統計量 $U = \sum U_i$

$$l_i = \frac{\partial \log f\left(Y_i, \beta\right)}{\partial \beta}$$

確率変数Yiの関数

スコア統計量の標本分布

- 確率変数Yi,i=1,・・・,nが互いに独立
  - Ui,i=1,・・・,nも互いに独立
  - 確率分布も同一となる

$$E[U]=0,\ \mathrm{var}[U]=\Im$$

多変量正規分布に漸近する(漸近標本分布)

$$U \sim N(0,\Im)$$

もとの標本では

$$U^T \Im^{-1} U \sim x^2(p)$$

対数尤度関数の近似

- 統計量の漸近的な標本分布を得る
  - テイラー級数展開

$$f(x) = f(t) + (x-t)\left[\frac{df}{dx}\right]_{x=t} + \frac{1}{2}(x-t)^2\left[\frac{d^2 f}{dx^2}\right]_{x=t} + \cdots$$

  - 単一パラメータを持つ対数尤度関数のパラメータ β の推定値b近傍での近似

$$l(\beta) = l(b) + (\beta - b)U(b) + \frac{1}{2}(\beta - b)^2 U'(b) + \cdots$$

$$\cong l(b) + (\beta - b)U(b) + \frac{1}{2}(\beta - b)^2 U'(b)$$

対数尤度関数の近似

- スコア関数(β =b)

$$U(b)=\frac{dl}{d\beta}$$

- スコア関数の一階微分の期待値による近似

$$U'=\frac{d^2l}{d\beta^2} \qquad E(U')=-\mathfrak{I}$$

$$l(\beta)\cong l(b)+(\beta-b)U(b)-\frac{1}{2}(\beta-b)^2\mathfrak{I}(b)$$

対数尤度関数の近似

- パラメータベクトル β の対数尤度関数の近似

$$l(\beta) \cong l(b) + (\beta - b) U(b) - \frac{1}{2}(\beta - b)^T \mathfrak{I}(b)(\beta - b)$$

- スコア関数のテイラー級数近似

$$U(\beta) = U(b) + (\beta - b) U'(b) + \frac{1}{2}(\beta - b)^2 U''(b) + \cdots$$

$$\cong U(b) + (\beta - b) U'(b)$$

U’を期待値E[U’]で近似→ $-\mathfrak{I}$

$$\cong U(b) - (\beta - b)\mathfrak{I}(b)$$

- ベクトルの場合 $U(\beta) \cong U(b) - \mathfrak{I}(b)(\beta - b)$

最尤推定量の標本分布

- 最尤推定量　$b = \hat{\beta}$
  - 対数尤度関数を最大にする　$U(b) = 0$

$$U(\beta) \cong U(b) - \mathfrak{I}(b)(\beta - b) = -\mathfrak{I}(b)(\beta - b)$$

  - 情報統計量が正則なら

$$\mathfrak{I}(b)^{-1} U(\beta) = -(\beta - b)$$

  - 情報統計量が一定ならスコア関数の期待値は0より

$$E\left[\mathfrak{I}(b)^{-1} U(\beta)\right] = \mathfrak{I}(b)^{-1} E\left[U(\beta)\right] = E\left[-(\beta - b)\right] = 0$$

bとβは一致する

最尤推定量の標本分布

- 情報行列 $\Im = E\left[UU^T\right]$
  - 対称性 $\left(\Im^{-1}\right)^T = \Im^{-1}$
  - 最尤推定量の分散共分散行列

$$E\left[(b-\beta)(b-\beta)^T\right] = \Im^{-1}E\left[UU^T\right]\Im^{-1} = \Im^{-1}$$

- パラメータ数p(自由度p)のカイ二乗分布

$$(b-\beta)^T\Im^{-1}(b)(b-\beta) \sim x^2(p)$$

  - ワルド統計量という

単一パラメータの時 $b \sim N\left(\beta, \Im^{-1}\right)$ と表せる

対数尤度比統計量

- モデルの適切さの評価
  - 飽和モデル(最大モデル，フルモデル) と比較
    - 推定されうる最大個数のパラメータを含む最も一般的なモデル
  - N個の観測値Yi,i=1,・・・,Nがあれば，N個の線形成分XiT β を対応させるN個のパラメータでモデルを表せる
    - 推定できるパラメータの数は，相異なる線形成分の数に等しい→繰り返し
      - 同じ線形成分(同じ共変量)をもつ観測値では，連続的な説明変数も同じになる

対数尤度比統計量

- 対象とするモデルは，飽和モデルと確率分布，連結関数が同じ一般化線形モデル
  - 飽和モデルのパラメータ数:m
  - 飽和モデルのパラメータベクトル: β max
  - パラメータベクトルの最尤推定量:bmax
    - 飽和モデルの尤度関数は最も大きいL(bmax;y)
    - 関心のあるモデルの尤度関数との比でモデルの適合度を評価:尤度比

$$\lambda = \frac{L\left(b_{\max}; y\right)}{L\left(b; y\right)}$$

対数尤度比統計量

- 尤度比の対数→対数尤度関数の差

$$\log\lambda = l\left(b_{\max}; y\right) - l\left(b; y\right)$$

  - $\log\lambda$ が大きいとモデルの適合が悪い
    - 標本分布を基に棄却域を決定
    - $2\log\lambda$:逸脱度
      - カイ二乗分布

逸脱度の標本分布

- 逸脱度→対数尤度比統計量:D

$$D = 2\left[l\left(b_{\max}; y\right) - l\left(b; y\right)\right]$$

- パラメータβの最尤推定量b　$U\left(b\right) = 0$
  - パラメータベクトルβの対数尤度関数の近似(テーラー展開)

$$l\left(\beta\right) \cong l\left(b\right) + \left(\beta - b\right) U\left(b\right) - \frac{1}{2}\left(\beta - b\right)^T \mathfrak{I}\left(b\right)\left(\beta - b\right)$$

$$l\left(\beta\right) - l\left(b\right) = -\frac{1}{2}\left(\beta - b\right)^T \mathfrak{I}\left(b\right)\left(\beta - b\right)$$

$$2\left[l\left(\beta\right) - l\left(b\right)\right] = -\left(\beta - b\right)^T \mathfrak{I}\left(b\right)\left(\beta - b\right) \sim x^2\left(p\right)$$

パラメータβの最尤推定量bはパラメータ数pのカイ二乗分布

逸脱度の標本分布

- 逸脱度D

$$D = 2\left[l\left(b_{\max}; y\right) - l\left(b; y\right)\right]$$

$$= 2\left[l\left(b_{\max}; y\right) - l\left(\beta_{\max}; y\right)\right]$$

$$- 2\left[l\left(b; y\right) - l\left(\beta; y\right)\right] + 2\left[l\left(\beta_{\max}; y\right) - l\left(\beta; y\right)\right]$$

飽和モデル:パラメータ数m:自由度mのカイ二乗分布

対象モデル:パラメータ数p:自由度pのカイ二乗分布

対象モデルが飽和モデルに適合すると零に近い正の数

逸脱度の標本分布の近似 $D \sim x^2\left(m - p, v\right)$非心パラメータv

自由度m-pのカイ二乗分布

逸脱度の標本分布

- 確率変数Yiが正規分布
  - 逸脱度Dはカイ二乗分布
    - ただし分散var[Yi]=σ2に依存
- 確率変数Yiが他の分布
  - 逸脱度はD近似的にカイ二乗分布
    - ちゃんと計算できるものも或る

仮説検定

- 検定の方法
  - ワルド統計量による検定
  - スコア統計量による検定
  - 逸脱度統計量の差を基にした検定
    - 単純化したモデルM0→帰無仮説H0
    - 一般的(複雑な)モデルM1→対立仮説H1
    - 検定を行うには，M0の線形成分がM1の線形成分の特別な場合となっている必要有

仮説検定

- モデルM0の帰無仮説H0

$$H_0 : \beta = \beta_0 = \left[\beta_1, \cdots, \beta_q\right]^t$$

- モデルM1の対立仮説H1

$$H_1 : \beta = \beta_1 = \left[\beta_1, \cdots, \beta_p\right]^t$$

  – ただしq<p<N

- 対立仮説H1に対する帰無仮説H0の検定

仮説検定

- 対立仮説H1に対する帰無仮説H0の検定
  - 逸脱度統計量の差

$$\Delta D = D_0 - D_1 = 2\left[l\left(b_{\max}; y\right) - l\left(b_0; y\right)\right] - 2\left[l\left(b_{\max}; y\right) - l\left(b_1; y\right)\right]$$
$$= 2\left[l\left(b_1; y\right) - l\left(b_0; y\right)\right]$$

    - 帰無仮説H0の支持

$$D_0 \sim x^2\left(N-q\right), D_1 \sim x^2\left(N-p\right) \Rightarrow \Delta D \sim x^2\left(p-q\right)$$

      - どちらのモデルもよく一致する→より簡単なM0を支持
    - 帰無仮説H0の棄却，対立仮説H1支持
      - ΔDが棄却域。カイ二乗分布上側100×αより大
      - M1の方がM0より有意によく記述している
        - » M1がよくあてはまるとは限らない