第 7 章　古典的回帰モデルの拡張 ― その 2

扱うトピック：**観測誤差の問題，操作変数，非正規性と大標本理論，大標本検定**

1　観測誤差の問題

理論モデル：　$y_i^* = \alpha + \beta x_i^* + \varepsilon_i$　（$i = 1, \cdots, n$）　(1)

観測値：　$x_i = x_i^* + u_i$,　$y_i = y_i^* + v_i$　(2)

ここで，u_i と v_i も，それぞれが平均 0 で独立な誤差項であり，x_i^*，y_i^* とは互いに独立であるとする．さらに，3 つの誤差項 ε_i，u_i，v_i は互いに独立であると仮定する．

説明変数 x^* が非確率的であっても，観測値 x は確率的となることに注意されたい．

LSE:

$$\hat{\beta} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \Big/ \sum_{i=1}^n (x_i - \bar{x})^2$$

$$= \sum_{i=1}^n (x_i^* - \bar{x}^* + u_i - \bar{u})(\alpha + \beta x_i^* + \varepsilon_i + v_i) \Big/ \sum_{i=1}^n (x_i - \bar{x})^2$$

$$= \beta \sum_{i=1}^n (x_i^* - \bar{x}^* + u_i - \bar{u}) x_i^* \Big/ \sum_{i=1}^n (x_i - \bar{x})^2$$

$$+ \sum_{i=1}^n (x_i^* - \bar{x}^* + u_i - \bar{u})(\varepsilon_i + v_i) \Big/ \sum_{i=1}^n (x_i - \bar{x})^2$$

（i）説明変数に観測誤差がない場合，すなわち，$x_i^* = x_i, u_i = 0$ の場合は，被説明変数に観測誤差があっても，$\hat{\beta}$ は不偏であり，一般には一致性をもつ．

（ii）説明変数に観測誤差がある場合には $\hat{\beta}$ は不偏でない．そして，

$$\operatorname*{plim}_{n \to \infty} \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = Q, \qquad 0 < Q < \infty \tag{3}$$

が成り立つならば，

$$\operatorname*{plim}_{n \to \infty} \hat{\beta} = \beta \left[1 - \frac{V(u_i)}{Q} \right] \tag{4}$$

となり，一致性が失われ，漸近的に下方へのバイアスをもたらす．

以上より，観測誤差の存在は，被説明変数については問題でないが，説明変数については重大であることがわかる．

その理由：(2) の関係を (1) に代入することにより，観測値に関するモデルが

$$y_i = \alpha + \beta x_i + \varepsilon_i - \beta u_i + v_i$$

のように表される．ここで，説明変数と誤差項は相関をもつ．実際，

$$Cov(x_i, \varepsilon_i - \beta u_i + v_i) = Cov(x_i^* + u_i, \varepsilon_i - \beta u_i + v_i) = -\beta V(u_i)$$

が成り立つ．このことが，$\hat{\beta}$ の一致性を失わせる原因となっている．

前章の第 5 節で説明した合理的期待モデルにおいても，合理的期待を観測値でおきかえて計算される LSE が一致性をもたない，ということを見たが，この事実は，形式上は観測誤差の問題とみなすことができる．

観測誤差が含まれる場合に適用される LSE の性質を，次の例題のように，コンピュータを使って調べることにしよう．

〔**例題** 7.1〕理論から導かれる回帰モデル

$$y_i^* = \alpha + \beta x_i^* + \varepsilon_i \qquad (i = 1,\cdots,10)$$

において，パラメータの値は $\alpha = 10$，$\beta = 0.8$ とする．また，実際の観測値は

$$x_i = x_i^* + u_i, \qquad y_i = y_i^* + v_i$$

で与えられるものとする．ここで，理論上の x_i^* は

i	1	2	3	4	5	6	7	8	9	10
x_i^*	1	1	2	2	3	3	4	4	5	5

であり，3 つの誤差項 ε_i，u_i，v_i は互いに独立に，平均 0，分散がそれぞれ $\sigma_\varepsilon^2 = 4$，$\sigma_u^2 = 4$，$\sigma_v^2 = 4$ の正規分布に従うものとする．このとき，β の LSE の分布をコンピュータ・シミュレーションにより求めよ．

（解）まず，互いに独立な正規乱数を使って ε_i，u_i，v_i を生成し，(1) および (2) から x_i と y_i を計算する．このようにして 10 個のデータを作り，β の LSE を求める．この手続きを何回か繰り返すことにより，推定量のヒストグラムができる．

⇒ **図 7-1:** 1,000 回の繰り返しにより得られたヒストグラムである．実線の曲線は，観測誤差がない場合の LSE の分布である．ヒストグラムは観測誤差がない場合の分布よりも左にシフトしており，明らかに下方へのバイアスをもたらしていることがわかる．実際，ヒストグラムの平均は 0.302 である．なお，標準偏差は 0.439 である．

2　操作変数推定量

$$y_i = \alpha + \beta x_i + w_i \tag{5}$$

において，説明変数 x_i は確率的であり，誤差項 w_i と相関をもつとする．ただし，説明変数は，

$$\operatorname*{plim}_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = Q, \qquad 0 < Q < \infty \tag{6}$$

となるような定常的なものであるとしよう．

今，説明変数 x_i とは相関があるが，誤差項 w_i とは無相関になる観測可能な変数 z_i が存在するならば，α と β の一致推定量 $\hat{\alpha}_{IV}$ と $\hat{\beta}_{IV}$ を次のように構成することができる．

$$\hat{\alpha}_{IV} = \bar{y} - \hat{\beta}_{IV}\bar{x}, \qquad \hat{\beta}_{IV} = \sum_{i=1}^{n}(z_i - \bar{z})(y_i - \bar{y}) \Big/ \sum_{i=1}^{n}(z_i - \bar{z})(x_i - \bar{x}) \tag{7}$$

これらを α と β の**操作変数**（IV: Instrumental Variables）**推定量**という．

IV 推定量は次のような方法で求められるものである．まず，(5) から次の 2 つの方程式が得られる．

$$\sum_{i=1}^{n} y_i = \alpha \sum_{i=1}^{n} 1 + \beta \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} w_i \tag{8}$$

$$\sum_{i=1}^{n} z_i y_i = \alpha \sum_{i=1}^{n} z_i + \beta \sum_{i=1}^{n} z_i x_i + \sum_{i=1}^{n} z_i w_i \tag{9}$$

このとき，$\hat{\alpha}_{IV}$ と $\hat{\beta}_{IV}$ は (8)，(9) の右辺の最後項を無視した方程式の解になる．このことは，IV 推定量が，LSE を求めるための正規方程式において，説明変数との積を作る際に代理の変数（今の例では z_i）を使って得られることを意味する．

一般に，誤差項と相関をもつような説明変数が含まれるモデルの推定において，その説明変数の代理として用いられる変数を**操作変数**といい，操作変数を使ってパラメータの推定量を求める方式を**操作変数法**という．IV 推定量が一致性をもつことは，(6) のほかに若干の追加的な条件のもとで，(7) の表現を変形することにより容易に示すことができる．

〔**例題** 7.2〕単回帰モデル

$$y_i = \alpha + \beta x_i + w_i, \qquad (i = 1, \cdots, 10)$$

において，説明変数 x_i と誤差項 w_i は相関をもつとする．これに対して，変数 z_i は，

$$x_i = z_i + w_i$$

の関係をみたし，w_i とは無相関である．z_i のデータが下のように与えられたとき，β の IV 推定量の分布をコンピュータ・シミュレーションにより求めよ．ただし，パラメータの値は $\alpha = 10$，$\beta = 0.8$ であり，誤差項 w_i は独立に，平均 0，分散がそれぞれ $\sigma_w^2 = 4$ の正規分布に従う．

i	1	2	3	4	5	6	7	8	9	10
z_i	1	1	2	2	3	3	4	4	5	5

（解）例題 7.1 と同様にして，正規乱数を使ってデータを作り出せばよい．

図 7-2： 1,000 回の繰り返しで得られた β の IV 推定量のヒストグラムである．実線の曲線は，**図 7-1** と同様に，観測誤差がない場合の LSE の分布である．ヒストグラムの平均は 0.992 であり，**図 7-1** の LSE のようなバイアスは見られないが，その代わりに標準偏差が1.315 と大きくなっている．平均 2 乗誤差（MSE）で比較すると，

MSE(LSE) $= (0.439)^2 + (0.302 - 0.8)^2 = 0.44$

MSE(IV) $= (1.315)^2 + (0.992 - 0.8)^2 = 1.77$

となり，LSE の方が良好なパフォーマンスを示すことになる．一致性は大標本の性質であるから，小標本のもとでは IV 推定量が LSE より劣ることがありうる

4　非正規性と大標本理論

回帰モデルに含まれるパラメータに関する推測，例えば，信頼区間を構成したり，t 検定や F 検定を厳密に行うことができる理由は，誤差項や推定量が正規分布に従う，という仮定があることによる．では，正規性が成立しない場合には，このような推測を

どのように行ったらよいであろうか．

今，単回帰モデル

$$y_i = \alpha + \beta x_i + \varepsilon_i \qquad (i = 1, \cdots, n) \tag{10}$$

において，誤差項 ε_i が互いに独立に平均 0，分散 σ^2 の同一分布（正規分布とは限らない）に従っているものとする．また，説明変数 x_i は確率的であり，あとで述べる条件をみたすものとする．

以下では，(10) のモデルから得られるさまざまな統計量の性質を大標本の場合について議論したい．大標本を前提にする理由は次の通りである．

（i）正規性の仮定がなければ，小標本における統計量の分布を導出することはほとんど不可能である．実際，β の LSE $\hat{\beta}$ を考えても，非正規性のもとでの分布がどうなるかは判定しがたい．このことから，小標本のもとでは，正規性の仮定がなければ有効な統計的推測が困難となる．

（ii）しかし，標本サイズが大きくなるにつれて，あとでみるように，統計量の分布が導出しやすくなり，したがって，統計的推測も可能となるからである．

さて，標本サイズ n に依存して定義される確率変数 Y_n の分布関数を $F_n(y)$ としよう．すなわち，$F_n(y) = P(Y_n \le y)$ とする．他方，確率変数 Y の分布関数を $F(y)$ とする．このとき，$F(y)$ のすべての連続点 y で $\lim_{n\to\infty} F_n(y) = F(y)$ となるならば，Y_n は Y に**分布収束**する，または Y_n は**極限（漸近）分布** F をもつという．分布収束あるいは極限分布の概念は，統計学において最も重要かつ美しい命題である次の定理（古典的な**中心極限定理**）の中にすでに現れている．

古典的な中心極限定理

$X_1, \cdots, X_n$ が互いに独立に平均 μ，分散 σ^2 の同一分布に従うとき，標本平均 $\overline{X}$ の分布は，n が大きくなるとともに，平均 μ，分散 σ^2/n の正規分布に近づく．すなわち，

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \to \quad N\left(\mu, \frac{\sigma^2}{n}\right)$$

あるいは，

$$\frac{\sqrt{n}\left(\overline{X} - \mu\right)}{\sigma} \quad \to \quad N(0,1)$$

となる．

この定理で重要なことは，分散が有限な $i.i.d.$ 確率変数は，いかなる分布に従っていようとも，その和の分布はもとの分布にかかわらず正規分布に近づくということである．すなわち，個々の分布が一様分布のような水平状の分布でも，また，指数分布のようなゆがんだ分布でも，さらに，二項分布のように離散的で一般に非対称な分布でも，それらの和の分布は対称で連続な釣り鐘状の分布に近づくということである．

ところで，β の LSE $\hat{\beta}$ は

$$\hat{\beta}=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}=\beta+\frac{\sum_{i=1}^{n}(x_i-\bar{x})\varepsilon_i}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \tag{11}$$

と表される．ε_i が正規分布に従わずに，しかも x_i が確率的ならば，$\hat{\beta}$ の漸近分布を導出することは古典的な中心極限定理では不可能である．

$\hat{\beta}$ の漸近分布を導出するための説明変数に関する追加的仮定：

(i) 定常

(ii) 誤差項とは独立

(iii) $$\operatorname*{plim}_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2=V(x_i)=Q, \qquad 0<Q<\infty \tag{12}$$

このとき，(11) から

$$\sqrt{n}\left(\hat{\beta}-\beta\right)=\frac{X_n}{Y_n}$$

を得る．ただし，

$$X_n=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(x_i-\bar{x})\varepsilon_i\,, \qquad Y_n=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

ここで，$\hat{\beta}$ の漸近分布だけではなく，一般に分布収束を考える際に有用な定理を以下にまとめておく．

漸近分布の導出に関する定理

(i) 確率変数 X_n が X に分布収束し，確率変数 Y_n が定数 c（$\neq 0$）に確率収束するとき，

(a) X_n+Y_n は $X+c$ に分布収束する．

(b) $X_n\times Y_n$ は $c\times X$ に分布収束する．

(c) X_n/Y_n は X/c に分布収束する．

(ii) 確率変数 X_n が X に分布収束し，$g(x)$ が連続関数であるとき，

$g(X_n)$ は $g(X)$ に分布収束する．

さて，(12) の仮定と上の定理の (i)-(c) から，$\sqrt{n}(\hat{\beta}-\beta)=X_n/Y_n$ の漸近分布は，X_n/Q の漸近分布に等しい．あとは X_n の漸近分布を考えればよい．ところで，

$$X_n=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(x_i-\bar{x})\varepsilon_i=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Z_i \tag{13}$$

において，$Z_i=(x_i-\bar{x})\varepsilon_i$ は独立ではなく，同一分布にも従わないが，仮定から

$$E\left(Z_i|Z_{i-1},\cdots,Z_1\right)=0 \tag{14}$$

が成立することがわかる．(14) の性質をみたす Z_i を，中心化された**マルチンゲール**

（Martingale），あるいは**マルチンゲール差**（Martingale Differences）という．中心化されたマルチンゲールに対しても，古典的な中心極限定理と同様の定理が成立する．

中心化されたマルチンゲールに対する中心極限定理

平均が 0 で有限な分散をもつ確率変数 Z_i に対して，

$$S_n = \sum_{i=1}^{n} Z_i\,,\quad T_n^2 = \sum_{i=1}^{n} Z_i^2\,,\quad s_n^2 = \sum_{i=1}^{n} E\left(Z_i^2\right) \tag{15}$$

を定義する．ここで，次の条件が成り立つものとする．

(a) $E\left(Z_i | Z_{i-1},\cdots,Z_1\right) = 0$

(b) $\underset{n\to\infty}{p\lim}\left(T_n^2 / s_n^2\right) = 1$

(c) 任意の正数 δ に対して $\lim_{n\to\infty}\left[\frac{1}{s_n^2}\sum_{i=1}^{n} E\left(Z_i^2 I\left(|Z_i| \geq \delta s_n\right)\right)\right] = 0$

ただし，(c) において，$I(A)$ は集合 A の定義関数を表す．このとき，

$$S_n/s_n = \sum\nolimits_{i=1}^{n} Z_i \bigg/ \sqrt{\sum\nolimits_{i=1}^{n} E\left(Z_i^2\right)} \quad \to \quad N(0,1)$$

が成立する．

中心化されたマルチンゲールは，分散が存在すれば必ず無相関である．このようなマルチンゲールに対する中心極限定理は，古典的な中心極限定理と違って，独立性や同一分布性を必要としない．その代わりに，条件の (b) にあるように，条件付き分散と無条件の分散の比が 1 に確率収束することを要請する．また，(c) は**リンデベルクの条件**と呼ばれ，分布のスソの厚さを制限するための条件である．もちろん，これらの条件は，平均 0，分散一定の独立，同一分布の場合には自動的にみたされる．

この定理を使って，(13) で定義された X_n の漸近分布を導出しよう．
$Z_i = (x_i - \bar{x})\varepsilon_i$ は中心化されたマルチンゲールであり，(15) は

$$S_n = \sum_{i=1}^{n}\left(x_i - \bar{x}\right)\varepsilon_i\,,\qquad T_n^2 = \sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 \varepsilon_i^2\,,\qquad s_n^2 = \sigma^2 \sum_{i=1}^{n} E\left[(x_i - \bar{x})^2\right]$$

となる．ここで，x_i が定常であることから，s_n^2 / n は $\sigma^2 Q$ に収束する．他方，x_i と ε_i の 4 次のモーメントが一様に有界ならば，T_n^2 / n は $\sigma^2 Q$ に確率収束することが示される．したがって，T_n^2 / s_n^2 は 1 に確率収束する．これが定理の条件 (b) である．さらに，このときには**リアプノフの条件**

$$\lim_{n\to\infty}\frac{1}{s_n^{2+\gamma}}\sum_{i=1}^{n} E\left[\left|(x_i - \bar{x})\varepsilon_i\right|^{2+\gamma}\right] = 0$$

が $\gamma = 2$ の場合に成立することから，リンデベルクの条件 (c) がみたされることになる．

以上より，(13) で定義された X_n の漸近分布が導かれ，この結果から β の LSE $\hat{\beta}$ の漸近分布が得られる．同様にして，α の LSE $\hat{\alpha}$ の漸近分布を導くこともできる．これらの結果を定理として述べることにしよう．

定理　単回帰モデル（10）において，次のことを仮定する．

（i）誤差項 ε_i は互いに独立に平均 0，分散 σ^2 の同一分布に従い，4 次のモーメントをもつ．

（ii）説明変数 x_i は平均 μ の定常過程に従い，誤差項と独立，一様に有界な 4 次のモーメントをもつ．

このとき，パラメータ α と β の LSE $\hat{\alpha}$ と $\hat{\beta}$ の分布について，次の中心極限定理が成り立つ．

$$\sqrt{n}(\hat{\alpha}-\alpha) \quad \to \quad N\left(0,\sigma^2\left(1+\mu^2/Q\right)\right), \qquad \sqrt{n}(\hat{\beta}-\beta) \quad \to \quad N\left(0,\sigma^2/Q\right)$$

〔**例題** 7.3〕単回帰モデル

$$y_i=\alpha+\beta x_i+\varepsilon_i\,, \qquad (i=1,\cdots,n)$$

において，誤差項 ε_i と説明変数 x_i は互いに独立で，ε_i は [-10,10] 上の一様分布，x_i は AR(1) モデル

$$x_i=m+\phi x_{i-1}+v_i\,, \qquad v_i \quad \sim \quad N(0,\sigma_v^2)$$

に従うものとする．$n=50$，$\alpha=10$，$\beta=0.8$，$m=4$，$\phi=0.6$，$\sigma_v^2=100$ のとき，α と β の LSE の分布をシミュレーションにより求め，漸近分布の結果と比較せよ．

（解）$\hat{\alpha}$ と $\hat{\beta}$ の分布をシミュレーションで求めるためには，まず，x_i のデータを生成する必要があるが，そのためには初期値 x_0 を指定しなければならない．ここでは，$x_0=\mu=m/(1-\phi)=10$ とした．そして，データが初期値に依存しないようにするために，80 個のデータを生成したのち，最初の 30 個を捨て去った．あとは，例題 7.1, 7.2 と同様にしてデータを作り出せばよい．漸近分布は，

$$\sigma^2=V(\varepsilon_i)=100/3, \quad \mu=m/(1-\phi)=10\,, \quad Q=V(x_i)=625/4$$

となることを使って，上の定理から

$$\hat{\alpha} \ \to \ N\left(\alpha,164/(3n)\right)=N(10,1.09)\,, \qquad \hat{\beta} \ \to \ N\left(\beta,16/(75n)\right)=N(0.8,0.0043)$$

となる．

図 7.3：1,000 回の繰り返しにより得られた $\hat{\alpha}$ の分布のヒストグラムと漸近分布を
図 7.4：$\hat{\beta}$ についての同様の図（標本サイズは 50）

マルチンゲール差に対する中心極限定理を使うことにより，推定量だけでなく，検定統計量についても漸近分布を導くことができる．例えば，検定問題

$$H_0{:}\beta=\beta_0 \qquad \text{vs} \qquad H_1{:}\beta\neq\beta_0$$

のための t 検定統計量

$$t_{\hat{\beta}}=\frac{\hat{\beta}-\beta_0}{\hat{\sigma}\Big/\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}}=\frac{\sqrt{n}(\hat{\beta}-\beta_0)}{\hat{\sigma}\Big/\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}}$$

を考えよう．ここで，

$$\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{\alpha} - \hat{\beta}x_i)^2$$

は σ^2 に確率収束することが示される．したがって，(12) の仮定，漸近分布の導出に関する定理の (1)-(c)，および LSE の漸近分布に関する一般的な定理を使うことにより，$t_{\hat{\beta}}$ の分布は，H_0 のもとで漸近的に標準正規分布に収束することがわかる．今の場合，t 検定統計量とはいっても，t 分布に従うわけではないことに注意されたい．

複数のパラメータに関する F 検定統計量についても，非正規性のもとでは，統計量が必ずしも F 分布に従うわけではない．そして，t 検定統計量と異なり，F 検定統計量そのものを使うことはできず，別の統計量を考える必要が出てくる．次の節では，このような場合に使われる大標本検定について説明する．

４　大標本検定

次の重回帰モデルを考えよう．

$$y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + \varepsilon_i \qquad (i = 1, \cdots, n) \tag{17}$$

ここで，誤差項 ε_i が非正規であったり，説明変数が確率的な場合には，線形制約仮説に関する F 検定統計量は，厳密には F 分布に従わない．また，たとえ古典的な仮定がみたされていても，制約仮説が非線形ならば，F 検定統計量は F 分布に従わない．そして，t 検定統計量が漸近的に標準正規分布をするのに対して，F 検定統計量の漸近分布は自明でなく，統計量としての有用性が低下してしまう．

そこで，このような場合には，F 検定統計量とは別の統計量を考える必要が出てくる．以下では，計量経済学において頻繁に使われる 3 つの大標本検定について説明することにしよう．

説明を簡単にするために，(17) は古典的な仮定をすべてみたす重回帰モデルであるとする．他方，帰無仮説としては，q 個の独立な制約仮説

$$H_0 : g(\theta) = c \tag{18}$$

を考える．ただし，$\theta = (\beta_0, \beta_1, \cdots, \beta_k, \sigma^2)'$ は $k+2$ 個の成分からなる列ベクトルであり，$g(\theta)$ は q 個の成分からなる列ベクトルで，各成分は微分可能な関数であるとする．

以上の設定のもとで，$y = (y_1, \cdots, y_n)$ は正規分布に従い，その同時密度関数は

$$f(y;\theta) = \frac{1}{\left(2\pi\sigma^2\right)^{n/2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_{1i} - \cdots - \beta_k x_{ki})^2\right]$$

となる．したがって，対数尤度関数は，

$$L(\theta) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_{1i} - \cdots - \beta_k x_{ki})^2 \tag{19}$$

で与えられる．

尤度比検定

(19) の対数尤度関数 $L(\theta)$ を最大にすることを考えてみよう．このとき，2 通りの最大化が考えられる．1 つは H_0 のもとでの最大化であり，最大値を与える θ を $\hat{\theta}_0$ としよう．もう 1 つは無条件の最大化であり，最大値を与える θ を $\hat{\theta}_1$ としよう．$\hat{\theta}_0$ と $\hat{\theta}_1$ は，それぞれ θ の制約付き MLE，無制約 MLE である．

ところで，対数尤度をこれらの MLE で評価すると，

$$L\left(\hat{\theta}_0\right) = -\frac{n}{2}\log(2\pi\hat{\sigma}_0^2) - \frac{n}{2}, \qquad L\left(\hat{\theta}_1\right) = -\frac{n}{2}\log(2\pi\hat{\sigma}_1^2) - \frac{n}{2}$$

と表すことができる．ただし，$\hat{\sigma}_0^2$ は σ^2 の H_0 のもとでの MLE，$\hat{\sigma}_1^2$ は無制約 MLE である．このとき，統計量

$$LR = -2\left(L\left(\hat{\theta}_0\right) - L\left(\hat{\theta}_1\right)\right) = n\log\frac{\hat{\sigma}_0^2}{\hat{\sigma}_1^2} \tag{20}$$

を**尤度比統計量**という．LR は Likelihood Ratio（尤度比）の略である．

LR は，H_0 のもとで漸近的に $\chi^2(q)$ に従うことが知られている．そして，LR 統計量の値が大きいときに H_0 を棄却する検定を**尤度比検定**という．LR が $\chi^2(q)$ の上側 100α% 点を越えるときに H_0 を棄却する検定は，漸近的に有意水準 α の検定となる．

尤度比検定を実行するためには，制約付き MLE と無制約 MLE の両方を計算する必要がある．この点で，尤度比検定は，以下で説明する他の 2 つの検定よりは少々面倒である．

ラグランジュ乗数検定

尤度比検定が対数尤度の値そのものを使うのに対して，ここで考える検定は，対数尤度の導関数を使うものである．$\hat{\theta}_0$ と $\hat{\theta}_1$ を，それぞれ θ の H_0 のもとでの MLE，無制約 MLE とする．ここで，$\hat{\theta}_1$ は $L(\theta)$ を無条件に最大化する値であるから，

$$\frac{\partial L(\hat{\theta}_1)}{\partial \theta} = 0$$

が恒等的に成り立っている．したがって，この導関数を統計量とすることはできない．他方，$\partial L(\hat{\theta}_0)/\partial\theta$ は一般に 0 とは異なる．また，中心極限定理により，

$$\frac{1}{\sqrt{n}}\frac{\partial L(\theta)}{\partial \theta} \quad \to \quad N\left(0, \Sigma(\theta)\right)$$

となることが知られている．ただし，$\Sigma(\theta)$ は $k+2$ 次の正方行列であり，その（i, j）要素 $\Sigma_{ij}(\theta)$ は，

$$\Sigma_{ij}(\theta) = -\lim_{n\to\infty} E\left(\frac{1}{n}\frac{\partial^2 L(\theta)}{\partial\theta_i \partial\theta_j}\right) \tag{21}$$

で与えられる．

以上のことから，検定統計量として，

$$LM = \frac{1}{n}\frac{\partial L\left(\hat{\theta}_0\right)}{\partial \theta'}\Sigma^{-1}\left(\hat{\theta}_0\right)\frac{\partial L\left(\hat{\theta}_0\right)}{\partial \theta} \tag{22}$$

を考えると，H_0 のもとで漸近的に $\chi^2(q)$ に従うことが示される．LM は**ラグラン**

ジュ乗数統計量と呼ばれ，LM の値が大きいときに H_0 を棄却する検定を**ラグランジュ乗数検定**という．

ラグランジュ乗数検定を行うためには，H_0 のもとでの MLE だけが必要であり，無制約 MLE は不要である．通常は，前者の方が後者よりは求めやすいので，ラグランジュ乗数検定は実行上の観点からは一番容易である．

ワルド検定

ここでは MLE に基づく検定を考えよう．$\hat{\theta}_1$ を無制約 MLE とするとき，

$$\sqrt{n}\left(\hat{\theta}_1-\theta\right) \quad \to \quad N\left(0,\Sigma^{-1}(\theta)\right)$$

となることが知られている．ただし，$\Sigma(\theta)$ は (21) で定義された $k+2$ 次の正方行列である．このとき，H_0 のもとで，

$$\sqrt{n}\left(g\left(\hat{\theta}_1\right)-c\right) \quad \to \quad N\left(0,\left(\partial g(\theta)\right)'\Sigma^{-1}(\theta)\partial g(\theta)\right)$$

が成り立つことも知られている．ただし，$\partial g(\theta)$ は，

$$\partial g(\theta)=\left(\frac{\partial g_1(\theta)}{\partial\theta},\cdots\cdots,\frac{\partial g_q(\theta)}{\partial\theta}\right)$$

で定義される $(k+2)\times q$ 行列である．

以上のことから，検定統計量

$$W=n\left(g(\hat{\theta}_1)-c\right)'\left[\left(\partial g(\hat{\theta}_1)\right)'\Sigma^{-1}(\hat{\theta}_1)\partial g(\hat{\theta}_1)\right]^{-1}\left(g(\hat{\theta}_1)-c\right) \qquad (23)$$

は，H_0 のもとで漸近的に $\chi^2(q)$ に従うことが示される．W を**ワルド統計量**と呼び，W が大きいときに H_0 を棄却する検定を**ワルド検定**という．

ワルド検定を行うためには，ラグランジュ乗数検定とは逆に，無制約 MLE だけが必要であり，H_0 のもとでの MLE は不要である．

以上，3 つの代表的な大標本検定について簡単に説明した．検定統計量の漸近分布は，帰無仮説のもとで，すべて制約数に等しい自由度をもつ χ^2 分布である．また，漸近的に同一の検出力をもたらすことも知られており，これら 3 つの検定は大標本においては同等である．

有限標本において，これら 3 つの検定の違いなどをより深く理解するために，次の例題を考えよう．

〔**例題** 7.4〕単純確率モデル

$$y_i=\mu+\varepsilon_i \qquad (i=1,\cdots,n)$$

を考える．ここで，誤差項 ε_i は互いに独立に平均 0，分散 σ^2 の正規分布に従う．このとき，検定問題

$$H_0{:}\,\mu=c \qquad \text{vs} \qquad H_1{:}\,\mu\neq c$$

に対する尤度比，ラグランジュ乗数，ワルド統計量を求め，それぞれの統計量の性質を調べよ．

（解）モデルに含まれるパラメータを $\theta=(\mu,\sigma^2)'$ とおく．(18) の制約条件は $g(\theta)=\mu$ であるから，$\partial g(\theta)=(1,0)'$ である．また，対数尤度は，

$$L(\theta)=-\frac{n}{2}\log(2\pi\sigma^2)-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2$$

で与えられる．このとき，H_0 のもとでの θ の MLE を $\hat{\theta}_0$，無制約 MLE を $\hat{\theta}_1$ とすると，

$$\hat{\mu}_0=c,\quad \hat{\sigma}_0^2=\frac{1}{n}\sum_{i=1}^{n}(y_i-c)^2,\quad \hat{\mu}_1=\bar{y},\quad \hat{\sigma}_1^2=\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2=\hat{\sigma}_0^2-(\bar{y}-c)^2$$

となることがわかる．したがって，

$$LR=n\log\frac{\hat{\sigma}_0^2}{\hat{\sigma}_1^2}=n\log\left[1+\frac{(\bar{y}-c)^2}{\hat{\sigma}_1^2}\right]=-n\log\left[1-\frac{(\bar{y}-c)^2}{\hat{\sigma}_0^2}\right]$$

となる．

次に，対数尤度を微分することにより，

$$\frac{\partial L(\theta)}{\partial\mu}=\frac{n(\bar{y}-\mu)}{\sigma^2},\qquad \frac{\partial L(\theta)}{\partial\sigma^2}=-\frac{n}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{i=1}^{n}(y_i-\mu)^2$$

$$\frac{\partial^2 L(\theta)}{\partial\mu^2}=-\frac{n}{\sigma^2},\quad \frac{\partial^2 L(\theta)}{\partial\mu\partial\sigma^2}=-\frac{n(\bar{y}-\mu)}{\sigma^4},\quad \frac{\partial^2 L(\theta)}{\partial(\sigma^2)^2}=\frac{n}{2\sigma^4}-\frac{1}{\sigma^6}\sum_{i=1}^{n}(y_i-\mu)^2$$

が得られる．このことから，

$$\frac{\partial L(\hat{\theta}_0)}{\partial\mu}=\frac{n(\bar{y}-c)}{\hat{\sigma}_0^2},\qquad \frac{\partial L(\hat{\theta}_0)}{\partial\sigma^2}=0,\qquad \Sigma(\theta)=\begin{pmatrix}1/\sigma^2 & 0\\ 0 & 1/(2\sigma^4)\end{pmatrix}$$

となり，

$$LM=\frac{1}{n}\frac{\partial L(\hat{\theta}_0)}{\partial\theta'}\Sigma^{-1}(\hat{\theta}_0)\frac{\partial L(\hat{\theta}_0)}{\partial\theta}=n(\bar{y}-c)^2\big/\hat{\sigma}_0^2$$

$$W=n\left(g(\hat{\theta}_1)-c\right)'\left[\left(\partial g(\hat{\theta}_1)\right)'\Sigma^{-1}(\hat{\theta}_1)\partial g(\hat{\theta}_1)\right]^{-1}\left(g(\hat{\theta}_1)-c\right)=n(\bar{y}-c)^2\big/\hat{\sigma}_1^2$$

が得られる．

以上より， 3 つの統計量はそれぞれ異なる値となるが，不等式

$$x<-\log(1-x),\qquad \log(1+x)<x$$

を利用することによって，大小関係

$$LM<LR<W$$

が常に成り立つことがわかる．もちろん，3 つの統計量は漸近的には H_0 のもとで $\chi^2(1)$ に従う．

上の例題では，3 つの統計量 LM，LR，W の間にアルファベット順の大小関係が成り立つことを，単純確率モデルの場合に示したが，このことは回帰モデルの場合にもあてはまることがわかっている．