

第 5 章: 重回歸分析 : OLS 漸近特性

Jeffrey Wooldridge (2018).

Introductory Econometrics: A Modern Approach
Seventh Edition. Cengage Learning.

2026-01-23

準備

必要なパッケージの読み込み

▶ wooldridge パッケージの読み込み

```
library(wooldridge)
```

5-1 一致性 (Consistency)

標本数と推定量

- ▶ 第 4 章までの性質（不偏性など）は、どのような標本サイズ n に対しても成立する**有限標本特性**（finite sample properties）であった。
- ▶ 第 5 章では、標本サイズ n が無限に大きくなったときの推定量の振る舞い、すなわち**漸近特性**（asymptotic properties）を扱う。
- ▶ 特に重要なのが、**一致性**（consistency）である。

OLS 推定量の一致性

- ▶ 標本サイズ n が大きくなるにつれて、推定量 $\hat{\beta}_j$ が真のパラメータ β_j に確率収束することを**一致性**という。

$$\text{plim}(\hat{\beta}_j) = \beta_j$$

- ▶ 仮定 MLR.1 (線形性) ~ MLR.4 (ゼロ条件付き平均) が満たされれば、OLS 推定量は**一致性**を持つ。
- ▶ 注: 仮定 MLR.6 (正規性) は**一致性**には不要である。

不偏性と一致性の違い

- ▶ **不偏性:** 期待値が真の値に等しいこと ($E(\hat{\beta}_j) = \beta_j$)。サンプルサイズが小さくても成立する。
- ▶ **一致性:** サンプルサイズが無限大の極限で、真の値に収束すること。
- ▶ OLS 推定量は、MLR.4 ($E(u|x) = 0$) が満たされれば不偏かつ一致推定量である。
- ▶ しかし、少し弱い仮定 ($E(u) = 0$ かつ $\text{Cov}(x, u) = 0$) の下では、不偏ではないが一致性を持つ場合がある。

5-2 漸近正規性と大規模標本での推論

漸近正規性 (Asymptotic Normality)

- ▶ 誤差項 u が正規分布に従わなくても (つまり MLR.6 が満たされなくても)、標本サイズ n が十分に大きければ、OLS 推定量 $\hat{\beta}_j$ の分布は近似的に正規分布に従う。

$$\sqrt{n}(\hat{\beta}_j - \beta_j) \xrightarrow{d} \text{Normal} \left(0, \frac{\sigma^2}{a_j^2} \right)$$

- ▶ これを**漸近正規性**と呼ぶ。
- ▶ これにより、大規模標本 (large sample) においては、誤差項の分布に関わらず t 検定や F 検定、信頼区間の構成が正当化される。
- ▶ a_j^2 は、説明変数 x_j の「他の説明変数と独立な部分の分散」を表す。

a_j^2 の意味

$$a_j^2 = \text{plim} \left(\frac{1}{n} \sum \hat{r}_{ij}^2 \right)$$

- ▶ x_j の「他の説明変数と独立な部分の分散」
- ▶ x_j が他の変数と強く相関していると、 a_j^2 は小さくなる。結果として、分散が大きくなる（精度悪化）

実証分析における含意

- ▶ 標本サイズが十分に大きい場合（例：数百以上）、データのヒストグラム（誤差項の分布）が正規分布に見えなくても、 t 値や p 値を通常通り解釈してよい。
- ▶ ただし、「十分に大きい」の基準はデータの性質による（極端な外れ値がある場合などは注意が必要）。

5-3 R によるシミュレーション

正規分布でない誤差項を持つモデル

- ▶ 誤差項が**一様分布**に従う場合のシミュレーションを行う。
- ▶ 真のモデル: $y = 1 + 2x + u$, $x \sim N(0, 1)$,
 $u \sim \text{Uniform}(-\sqrt{3}, \sqrt{3})$ (分散は 1)

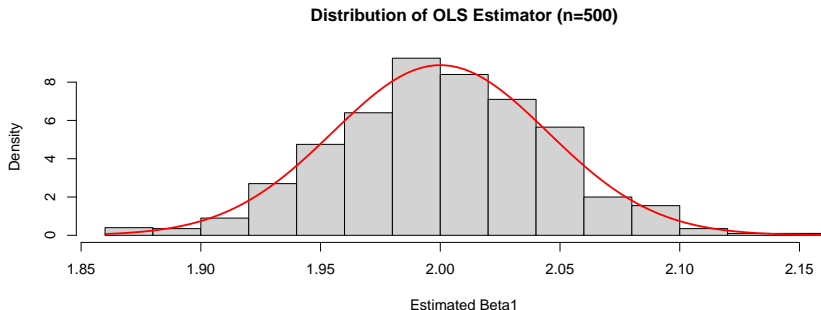
```
set.seed(123)
n_sim <- 1000 # シミュレーション回数
beta1_estimates <- numeric(n_sim)
n <- 500      # サンプルサイズ (比較的大きい)

for(i in 1:n_sim){
  x <- rnorm(n)
  u <- runif(n, min = -sqrt(3), max = sqrt(3)) # 非正規
  y <- 1 + 2 * x + u
  beta1_estimates[i] <- coef(lm(y ~ x))[2]
}
```

推定量の分布の確認

ヒストグラムと正規分布曲線の重ね書き

```
hist(beta1_estimates, probability = TRUE,  
     main = "Distribution of OLS Estimator (n=500)",  
     xlab = "Estimated Beta1")  
curve(dnorm(x, mean = 2, sd = sd(beta1_estimates)),  
     add = TRUE, col = "red", lwd = 2)
```



5-4 大規模標本でのラグランジュ乗数検定 (LM 検定)

LM 検定 (Lagrange Multiplier Statistic)

- ▶ 複数の制約に関する検定として、F 検定の代わりに使用されることがある。
- ▶ 特に、**制約付きモデル (Restricted Model)** のみを推定すればよいため、制約なしモデルの推定が困難な場合に便利である。
- ▶ 大規模標本において、LM 統計量はカイ二乗分布 χ_q^2 に従う (q は制約の数)。

LM 検定の手順

1. 制約付きモデル (H_0 を課したモデル) を推定し、残差 \tilde{u} を保存する。
2. 残差 \tilde{u} を、制約なしモデルの**すべての**独立変数で回帰する (補助回帰)。
3. この補助回帰の決定係数 $R_{\tilde{u}}^2$ を取得する。
4. $LM = n \cdot R_{\tilde{u}}^2$ を計算する。
5. 自由度 q のカイ二乗分布と比較して p 値を求める。

LM 検定の例 (crime1)

- ▶ $H_0 : \beta_{\text{avgsen}} = 0, \beta_{\text{tottime}} = 0$ (刑罰の厳しさは犯罪率に影響しない)

```
data(crime1)
n <- nrow(crime1)
# 1. 制約付きモデル (刑罰変数なし)
model_r <- lm(narr86 ~ pcnv + ptime86 + qemp86,
              data = crime1)
```

LM 検定の例 (crime1): 続き

```
model_r

##
## Call:
## lm(formula = narr86 ~ pcnv + ptime86 + qemp86, data = crime1)
##
## Coefficients:
## (Intercept)          pcnv          ptime86          qemp86
##    0.71177    -0.14993    -0.03442    -0.10411

# 残差の保存
u_tilde <- resid(model_r)
```

LM 検定の例 (crime1): 続き

```
# 2. 補助回帰 (残差をすべての変数で回帰)
```

```
# avgssen, tottime を追加
```

```
model_aux <- lm(u_tilde ~ pcnv + ptime86 + qemp86 +  
                avgssen + tottime, data = crime1)
```

```
model_aux
```

```
##
```

```
## Call:
```

```
## lm(formula = u_tilde ~ pcnv + ptime86 + qemp86 + avgssen + tottime,  
##     data = crime1)
```

```
##
```

```
## Coefficients:
```

```
## (Intercept)          pcnv          ptime86          qemp86          avgssen  
## -0.005711      -0.001297      -0.004839      0.001022      -  
0.007049      0.012095
```

LM 検定の計算

```
# 3. 決定係数の取得
```

```
summary(model_aux)$r.squared
```

```
## [1] 0.001493846
```

```
n
```

```
## [1] 2725
```

```
# 4. LM 統計量の計算
```

```
lm_stat <- n * summary(model_aux)$r.squared
```

```
lm_stat
```

```
## [1] 4.070729
```

LM 検定の判定

```
# 5. p 値の計算 (自由度  $q = 2$ )  
p_value <- 1 - pchisq(lm_stat, df = 2)  
p_value
```

```
## [1] 0.1306328
```

- ▶ p 値が 0.05 より大きければ、帰無仮説を棄却できない。
- ▶ F 検定の結果と (漸近的に) 類似するはずである。

まとめ

まとめ

- ▶ OLS 推定量は、標本サイズが大きくなると真の値に確率収束する（**一貫性**）。
- ▶ 誤差項が正規分布でなくても、標本サイズが十分に大きければ、OLS 推定量は近似的に正規分布に従う（**漸近正規性**）。
- ▶ これにより、大規模標本では通常の t 検定や F 検定が妥当となる。
- ▶ LM 検定は、大規模標本において F 検定の代替として利用できる（補助回帰を用いる）。