漸近理論（大標本理論）攻略マニュアル

計量経済学 基礎からの解説シリーズ 第4部

目次

1　なぜ漸近理論が必要か：全体の地図　4

1.1　有限標本の正確な分布とその限界 4

1.2　発想の転換：N を大きくしたときの振る舞いを調べる 4

1.3　漸近理論が答える2つの問い 4

1.4　記号の約束：確率変数の列 5

1.5　全体を貫く3段階の証明の型 5

1.6　この教材の地図 5

2　確率収束と大数の法則　7

2.1　ふつうの数列の収束（復習） 7

2.2　確率収束の定義 7

2.3　マルコフの不等式とチェビシェフの不等式 8

2.4　大数の法則 9

2.5　図で見る大数の法則 10

2.6　一致性：推定量が真の値に近づくこと 11

2.7　連続写像定理 12

3　OLS の一致性　14

3.1　ステップ1：推定量を分解する 14

3.2　ステップ2：分子の極限 15

3.3　ステップ3：分母の極限と結論 15

3.4 内生性と欠落変数バイアス 16
3.5 行列版：一般の重回帰 16

4 分布収束と中心極限定理 18
4.1 なぜ $\sqrt{N}$ をかけるのか 18
4.2 分布収束の定義 19
4.3 確率収束との関係 19
4.4 中心極限定理 20
4.5 図で見る中心極限定理 20
4.6 ベクトル版の CLT 21
4.7 連続写像定理（分布収束版） 22

5 Slutsky の定理と OLS の漸近正規性 23
5.1 Slutsky の定理 23
5.2 特に重要な2つの帰結 24
5.3 OLS の漸近正規性（単回帰） 24
5.4 均一分散の場合 26
5.5 行列版 26

6 漸近分散・標準誤差・t 統計量 28
6.1 漸近分散と標準誤差 28
6.2 t 統計量が標準正規分布に近づく理由 28
6.3 頑健な標準誤差が必要な理由 30

7 ベクトルの場合：線形変換・二次形式・Wald 検定 32
7.1 多変量正規分布の2つの性質 32
7.2 線形変換 32
7.3 二次形式 33
7.4 Wald 検定 34

8 デルタ法 36
8.1 デルタ法の主張 36
8.2 直感：接線で近似する 37

8.3 証明 38
8.4 ベクトル版 38

9 o_p・O_p 記法と周辺の用語 40
9.1 定義 40
9.2 例 41
9.3 計算規則 42
9.4 その他の用語 43

10 総まとめと練習問題 45
10.1 道具箱の一覧 45
10.2 OLS の推論の各ステップと使う道具 46
10.3 練習問題 46
10.4 解答 47

第1章

なぜ漸近理論が必要か：全体の地図

> **ここまでの流れ**　学部の回帰分析では、誤差項が正規分布に従うと仮定する（MLR.6）ことで、t 統計量の「正確な分布」を導きました。
> **この章の問い**　誤差の正規性を仮定できないとき、推定量の性質や検定をどう扱えばよいのでしょうか。
> **この章で分かること**　「$N \to \infty$ の極限で近似する」という発想と、この教材全体を貫く3段階の証明の型（分解 → 大数の法則・中心極限定理 → 連続写像定理・Slutsky の定理）です。

1.1　有限標本の正確な分布とその限界

学部の回帰分析では、誤差項 u_i が正規分布に従うと仮定すると、t 統計量が自由度 $N-K$ の t 分布にぴったり従うことを示せました。ここで「ぴったり」とは、サンプルサイズ N がどんなに小さくても成り立つという意味で、**有限標本の正確な分布**と呼ばれます。

しかし現実のデータで、誤差項が正規分布に従う保証はどこにもありません。さらに困ったことに、正規性を外してしまうと、推定量の正確な分布は**ほとんどの場合求められません**。推定量は $\hat{\beta} = \beta + (X^\top X)^{-1} X^\top u$ のように、誤差の複雑な関数だからです。

1.2　発想の転換：N を大きくしたときの振る舞いを調べる

そこで発想を変えます。正確な分布を求めることを諦め、**サンプルサイズ N を限りなく大きくしたとき、推定量がどう振る舞うか**を調べます。得られた極限の姿を、「N が十分大きいときの近似」として使います。この考え方が**漸近理論**（大標本理論）です。「漸近」とは「$N \to \infty$ に近づける」という意味です。

1.3　漸近理論が答える2つの問い

漸近理論で答えたい問いは2つあります。

1. $N \to \infty$ で、推定量は真の値に近づくか？　⇒ **確率収束・大数の法則・一致性**（第2章・第3章）
2. そのとき、推定量と真の値のずれは、どんな分布で近似できるか？　⇒ **分布収束・中心極限定理・漸近正規性**（第4章以降）

1つ目は「推定量は的に当たるか」、2つ目は「当たらないときのズレ方はどんな形か」という問いです。1つ目だけでは、推定量の良し悪しは分かっても、信頼区間や検定は作れません。ズレの分布が分かって初めて、t 検定や F 検定（漸近的には Wald 検定）ができます。

1.4 記号の約束：確率変数の列

以下では、x_N のように添字 N のついたものは、「サンプルサイズ N のデータから計算した確率変数」を表します。$N = 1, 2, 3, \ldots$ と増やしていくと、確率変数の列 $x_1, x_2, x_3, \ldots$ ができます。漸近理論は、この列の「行き着く先」を調べる理論です。

例えば標本平均 $\bar{w}_N = \frac{1}{N}\sum_{i=1}^{N} w_i$ は、N ごとに異なる確率変数で、$\bar{w}_1, \bar{w}_2, \bar{w}_3, \ldots$ という列をなします。

1.5 全体を貫く3段階の証明の型

この教材で扱う証明は、ほとんどすべて次の3段階でできています。

要点：証明の3段階

1. **分解する**：推定量を「真の値 $+$（標本平均の形をした項）」に書き直します（学部で見た $\hat{\beta} = \beta + \ldots$ の分解です）。
2. **標本平均の部分に法則を当てはめる**：標本平均の部分に、大数の法則（確率収束）または中心極限定理（分布収束）を使います。
3. **部品を組み立てる**：連続写像定理と Slutsky の定理で、各部品の極限を合成します。

1.6 この教材の地図

章	内容	答える問い
2	確率収束・大数の法則・一致性・連続写像定理	推定量は真の値に近づくか
3	OLS の一致性	OLS はいつ一致するか
4	分布収束・中心極限定理	ズレはどんな分布に近づくか
5	Slutsky の定理・OLS の漸近正規性	OLS のズレの分布は何か
6	漸近分散・標準誤差・t 統計量	なぜ $\lvert t\rvert > 1.96$ で棄却できるか
7	二次形式・Wald 検定	複数の制約をまとめて検定するには
8	デルタ法	非線形関数の推定量の分布は
9	o_p・O_p 記法と周辺の用語	証明を簡潔に書く道具
10	総まとめと練習問題	身についたかの確認

要点

- 誤差の正規性を仮定しないと、推定量の正確な分布は通常求められないため、$N \to \infty$ の極限で近似します。
- 漸近理論は、「推定量は真の値に近づくか」（確率収束）と「ズレはどんな分布か」（分布収束）の2つに答えます。
- 証明は「分解 → 大数の法則・中心極限定理 → 連続写像定理・Slutsky の定理」の3段階が基本の型です。

次の章へ　まず1つ目の問い、「推定量は真の値に近づくか」から始めます。確率変数の列が定数に近づくとはどういうことかを、確率収束として定義します。

大学院とのつながり　大学院の講義では、GLS・2SLS・GMM のように不偏性を示しにくい推定量を多く扱います。そのため、不偏性よりも一致性と漸近分布が主役になります。

第2章

確率収束と大数の法則

ここまでの流れ　漸近理論は「$N \to \infty$ で推定量がどうなるか」を調べる理論で、1つ目の問いは「推定量は真の値に近づくか」でした。
この章の問い　データによって値が変わる確率変数の列が、定数に「近づく」とはどういう意味でしょうか。標本平均は母集団平均に近づくのでしょうか。
この章で分かること　確率収束の定義、チェビシェフの不等式を使った大数の法則の証明、一致性と不偏性の違い、そして収束を組み合わせる道具である連続写像定理です。

2.1　ふつうの数列の収束（復習）

確率変数を扱う前に、ふつうの数列 a_N の収束を思い出します。$a_N \to a$ とは、次のことでした。

> 任意の $\varepsilon > 0$ に対して、ある N_0 があって、$N \geq N_0$ ならば $|a_N - a| < \varepsilon$ が成り立つ。

「どんなに狭い幅 ε を指定されても、十分先の項はすべてその幅に入る」という意味です。例えば $a_N = 1 + 1/N$ は 1 に収束します。幅 ε を指定されたら、$N_0 > 1/\varepsilon$ ととればよいからです。

確率変数 x_N では、この定義はそのままでは使えません。x_N はデータによって値が変わるので、「必ず幅に入る」とは言えないからです。そこで「幅に入らない確率」が 0 に近づくことを要求します。

2.2　確率収束の定義

要点：確率収束

任意の $\varepsilon > 0$ について

$$\lim_{N \to \infty} \Pr\bigl(|x_N - a| > \varepsilon\bigr) = 0 \tag{2.1}$$

が成り立つとき、x_N は a に**確率収束**するといい、$x_N \xrightarrow{p} a$ または $\operatorname{plim} x_N = a$ と書きます。ベクトルや行列の場合は、すべての要素が確率収束すること（同値な言い方：$\|x_N - a\| \xrightarrow{p} 0$）と定義します。

数式の読み方

- $p_N(\varepsilon) = \Pr(|x_N - a| > \varepsilon)$ は、ε を固定すればただの数列です。確率収束とは、この数列が（どんな $\varepsilon > 0$ についても）0 に収束することです。
- ε は「許容する誤差の幅」です。ε を小さくするほど厳しい要求になりますが、どんなに小さくても、N を大きくすれば外れる確率はいくらでも 0 に近くなる、というのが確率収束です。

補足：つまずきポイント

確率収束は「N が大きければ x_N は**必ず** a の近くにある」という意味ではありません。どんなに N が大きくても、運悪く大きくずれる可能性は残ります。その可能性（確率）が 0 に向かう、というのが正確な意味です。

2.3 マルコフの不等式とチェビシェフの不等式

確率収束を示すための基本道具が、次の2つの不等式です。

要点：マルコフの不等式

任意の確率変数 x、$\varepsilon > 0$、$p > 0$ について

$$\Pr(|x| \geq \varepsilon) \leq \frac{\mathrm{E}(|x|^p)}{\varepsilon^p}. \tag{2.2}$$

証明は指示関数 $\mathbf{1}(\cdot)$（条件が成り立てば 1、そうでなければ 0）を使うと1行ずつ追えます。$Z = |x|^p$ (≥ 0)、$c = \varepsilon^p$ とおきます。

$$\begin{aligned}
c \cdot \mathbf{1}(Z \geq c) &\leq Z && (Z \geq c \text{ なら左辺は } c \leq Z,\ Z < c \text{ なら左辺は } 0 \leq Z) \\
\mathrm{E}\big[c \cdot \mathbf{1}(Z \geq c)\big] &\leq \mathrm{E}(Z) && (\text{両辺の期待値をとる}) \\
c \Pr(Z \geq c) &\leq \mathrm{E}(Z) && (\mathrm{E}[\mathbf{1}(Z \geq c)] = \Pr(Z \geq c)) \\
\Pr(|x|^p \geq \varepsilon^p) &\leq \frac{\mathrm{E}(|x|^p)}{\varepsilon^p} && (Z, c \text{ を元に戻す})
\end{aligned}$$

$\{|x| \geq \varepsilon\} = \{|x|^p \geq \varepsilon^p\}$ なので、結論が得られます。

要点：チェビシェフの不等式

分散が有限な確率変数 Z について

$$\Pr\big(|Z - \mathrm{E}Z| \geq \varepsilon\big) \leq \frac{\mathrm{Var}(Z)}{\varepsilon^2}. \tag{2.3}$$

マルコフの不等式で $x = Z - \mathrm{E}Z$、$p = 2$ とおくと、$\mathrm{E}(|x|^2) = \mathrm{Var}(Z)$ なので直ちに得られます。

数式の読み方

「分散が小さいものは、平均から大きく外れにくい」ことを定量的に述べた不等式です。分散を 0 に近づけられれば、外れる確率も 0 に近づきます。これが確率収束を示す最も基本的な方法です。

計算例：チェビシェフの不等式の使い方

Z の分散が 9 のとき、平均から 0.5 以上外れる確率は $9/0.5^2 = 36$ 以下、つまりこの不等式は何も教えてくれません（確率は 1 以下です）。ところが、Z が $N = 400$ 個の独立な観測の平均で、1 つ当たりの分散が 9 なら $\mathrm{Var}(Z) = 9/400$ となり、上限は $\frac{9/400}{0.25} = 0.09$ になります。観測数を増やして分散を下げれば、不等式が意味を持つようになります。

2.4 大数の法則

要点：大数の弱法則（WLLN）

$w_1, \ldots, w_N$ が IID で $\mathrm{E}|w_i| < \infty$ なら

$$\bar{w}_N = \frac{1}{N}\sum_{i=1}^{N} w_i \xrightarrow{p} \mathrm{E}(w_i). \tag{2.4}$$

「標本平均は母集団平均に確率収束する」。

証明は、分散も有限という、やや強い仮定のもとで示します。$\mu = \mathrm{E}(w_i)$、$\sigma^2 = \mathrm{Var}(w_i) < \infty$ とします。まず標本平均の平均と分散を計算します。

$$\begin{aligned}
\mathrm{E}(\bar{w}_N) &= \frac{1}{N}\sum_{i=1}^{N}\mathrm{E}(w_i) = \mu && \text{（期待値の線形性）} \\
\mathrm{Var}(\bar{w}_N) &= \frac{1}{N^2}\sum_{i=1}^{N}\mathrm{Var}(w_i) && \text{（}w_i\text{ が独立なので分散は和の分散に等しい）} \\
&= \frac{N\sigma^2}{N^2} = \frac{\sigma^2}{N} && \text{（}\mathrm{Var}(w_i) = \sigma^2\text{ を代入）}
\end{aligned}$$

次に、チェビシェフの不等式 (2.3) を $Z = \bar{w}_N$ に適用します。任意の $\varepsilon > 0$ について

$$\begin{aligned} \Pr\bigl(|\bar{w}_N - \mu| \geq \varepsilon\bigr) &\leq \frac{\mathrm{Var}(\bar{w}_N)}{\varepsilon^2} && \text{（チェビシェフの不等式）} \\ &= \frac{\sigma^2}{N\varepsilon^2} \xrightarrow{N\to\infty} 0 && \text{（}\mathrm{Var}(\bar{w}_N) = \sigma^2/N \text{ を代入）} \end{aligned}$$

となり、確率収束 (2.1) が示せました。

補足

定理そのものは $\mathrm{E}|w_i| < \infty$ だけで成り立ちますが、その証明には別の技術（打ち切り法など）が必要なので省略します。以下で使う場面では、上の証明で十分です。

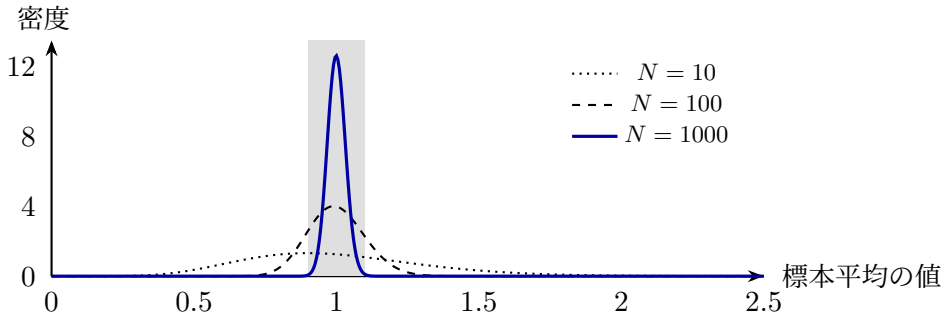
2.5 図で見る大数の法則

標本平均の分布が N とともにどう変わるかを、正確な分布が分かる例で確かめます。w_i を平均 1 の指数分布からの IID 標本とすると、$N\bar{w}_N = \sum w_i$ はガンマ分布に従い、$\bar{w}_N$ の密度は

$$f_N(x) = \frac{N^N x^{N-1} e^{-Nx}}{\Gamma(N)}, \qquad x > 0 \tag{2.5}$$

と書けます。

図 2.1　平均 1 の指数分布からの標本平均 $\bar{w}_N$ の密度



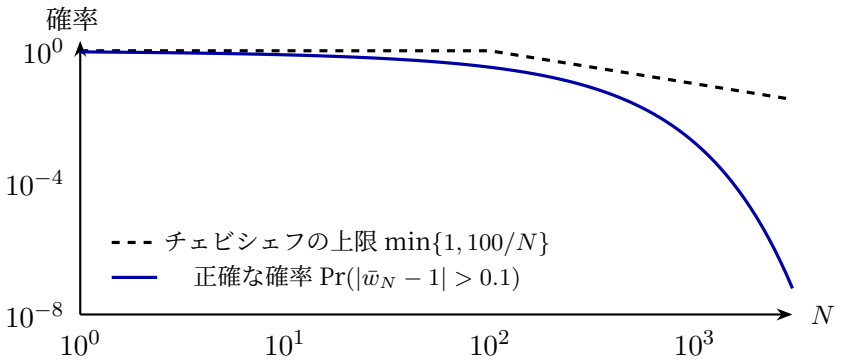
$f_N(x) = N^N x^{N-1} e^{-Nx}/\Gamma(N)$（ガンマ分布の密度）を描いた。灰色の帯は $[1-0.1, 1+0.1]$（$\varepsilon = 0.1$）。

図の読み方

1. 3本の曲線は $N = 10, 100, 1000$ のときの $\bar{w}_N$ の分布で、真の平均は 1 です。
2. 灰色の帯は $[1-\varepsilon, 1+\varepsilon]$（$\varepsilon = 0.1$）です。帯の外側の面積が $\Pr(|\bar{w}_N - 1| > 0.1)$ にあたります。
3. N が大きくなるほど分布は帯の中に集中し、帯の外側の面積は 0 に近づきます。これが確率収束を絵にしたものです。

次に、帯の外に出る確率の大きさを具体的に見ます。図 2.2 は、$\varepsilon = 0.1$、$\sigma^2 = 1$ のときの、正確な確率とチェビシェフの上限 $\sigma^2/(N\varepsilon^2) = 100/N$ を比べたものです。

図 2.2　帯の外に出る確率：正確な値とチェビシェフの上限



正確な確率は $\bar{w}_N \sim \text{Gamma}(N, \text{scale} = 1/N)$ の分布関数から計算した。10^{-9} 未満は 10^{-9} として描いている。

図の読み方

1. 縦軸・横軸とも対数目盛です。青の実線が正確な確率、黒の破線がチェビシェフの上限です。
2. 上限は、つねに正確な確率より上にあります（上限なので当然です）。
3. 上限は粗い見積もりで、正確な確率の方がずっと速く 0 に近づきます。しかし、上限が 0 に向かうだけで確率収束を示すには十分です。確率収束の証明に必要なのは「0 に行くこと」であり、「0 への行き方の速さ」ではありません。

2.6　一致性：推定量が真の値に近づくこと

要点：一致推定量

推定量 $\hat{\theta}_N$ が $\hat{\theta}_N \xrightarrow{p} \theta$（真の値）を満たすとき、$\hat{\theta}_N$ は θ の**一致推定量**であるといいます。

大数の法則は「標本平均は母平均の一致推定量である」と言い換えられます。

不偏性（$\mathrm{E}\hat{\theta}_N = \theta$）は「**固定された N で**、平均的に当たる」という性質です。一致性は「N を増やせば、当たる確率が 1 に近づく」という性質です。この2つは、互いにどちらも他方を含意しません。

計算例：不偏性と一致性が別物である例（w_i は IID、$\mathrm{E}w_i = \mu$、$\mathrm{Var}w_i = \sigma^2 > 0$）

不偏だが一致的でない：$\hat{\theta}_N = w_1$（1人目のデータだけを使う）を考えます。$\mathrm{E}(w_1) = \mu$ なので不偏です。しかし N が増えても w_1 の分布は変わらないので、$\Pr(|w_1 - \mu| > \varepsilon)$ は 0 に近づきません。

一致的だが不偏でない：$\tilde{\sigma}_N^2 = \frac{1}{N}\sum(w_i - \bar{w})^2$ を考えます。$\mathrm{E}(\tilde{\sigma}_N^2) = \frac{N-1}{N}\sigma^2 \neq \sigma^2$ なので偏りがありますが、$N \to \infty$ で偏りは消えて $\tilde{\sigma}_N^2 \xrightarrow{p} \sigma^2$ となります。

補足

大学院の講義では、不偏性を示すのが難しい推定量（GLS、2SLS、GMM など）を多く扱います。そのため、主に一致性を使います。

2.7 連続写像定理

要点：連続写像定理（確率収束版）

関数 g が点 c で連続であり、$x_N \xrightarrow{p} c$ なら

$$g(x_N) \xrightarrow{p} g(c). \tag{2.6}$$

ベクトル・行列値の x_N、g でも同様です。

証明は連続性の定義をそのまま使います。任意の $\varepsilon > 0$ をとります。g が c で連続なので、ある $\delta > 0$ があって

$$|x - c| < \delta \;\Rightarrow\; |g(x) - g(c)| < \varepsilon$$

が成り立ちます。この対偶は「$|g(x) - g(c)| \geq \varepsilon \Rightarrow |x - c| \geq \delta$」です。したがって

$$\begin{aligned} \Pr\big(|g(x_N) - g(c)| \geq \varepsilon\big) &\leq \Pr\big(|x_N - c| \geq \delta\big) && \text{（対偶より、左の事象は右の事象に含まれる）} \\ &\to 0 && \text{（}x_N \xrightarrow{p} c\text{ の定義）} \end{aligned}$$

となります。

数式の読み方：よく使う形

$a_N \xrightarrow{p} a$、$b_N \xrightarrow{p} b$ なら

$$a_N + b_N \xrightarrow{p} a + b, \qquad a_N b_N \xrightarrow{p} ab, \qquad a_N / b_N \xrightarrow{p} a/b \ \ (b \neq 0).$$

(a_N, b_N) を1つのベクトルとみて、連続関数 $g(a,b) = a+b,\ ab,\ a/b$ に連続写像定理を使っています。行列でも同様で、$Z_N \xrightarrow{p} A$ かつ A が正則なら $Z_N^{-1} \xrightarrow{p} A^{-1}$ です。逆行列の各要素は、行列式が 0 でない点の近くで元の行列の要素の連続関数だからです（2×2 の逆行列の公式を思い出してください）。

補足：つまずきポイント

条件「$b \neq 0$」「A が正則」は本質的です。極限で割り算や逆行列が定義できない点では g が連続でないので、定理は使えません。OLS の一致性でランク条件（$\mathrm{E}(x_i^\top x_i)$ が正則）が必要なのは、まさにこのためです。

補足：「Slutsky の定理」という名前について

上の連続写像定理を「Slutsky の定理」と呼ぶ文献もあります。多くの教科書では、「Slutsky の定理」は第5章で扱う「分布収束と確率収束の組み合わせ」の定理を指すことが多く、名前の使い方に流儀の違いがあります。中身はどちらも使います。

要点

- 確率収束 $x_N \xrightarrow{p} a$ は、「a から ε 以上ずれる確率が、$N \to \infty$ で 0 に近づく」ことです。必ず近いのではなく、ずれる確率が消えるのです。
- 大数の弱法則：IID 標本の標本平均は母平均に確率収束します。分散が有限なら、チェビシェフの不等式から $\Pr(|\bar{w}_N - \mu| \geq \varepsilon) \leq \sigma^2/(N\varepsilon^2)$ で証明できます。
- 一致性（$N \to \infty$ で当たる）と不偏性（固定 N で平均的に当たる）は別の性質です。大学院では主に一致性を使います。
- 連続写像定理により、確率収束は和・積・商・逆行列に引き継がれます。ただし、極限で割り算・逆行列が定義できることが条件です。

次の章へ　大数の法則と連続写像定理は、すでに OLS の一致性を示すのに十分な道具です。次の章で、第1章で見た3段階の型を実際に使ってみます。

大学院とのつながり　行列版の連続写像定理（逆行列の収束）は、OLS・GLS・2SLS・GMM の一致性の証明に毎回登場します。

第3章

OLS の一致性

ここまでの流れ　確率収束・大数の法則・連続写像定理という3つの道具がそろいました。
この章の問い　OLS 推定量は、いつ真の係数に確率収束するのでしょうか。内生性（誤差と説明変数の相関）があると何が起こるのでしょうか。
この章で分かること　「分解 → 大数の法則 → 連続写像定理」の3段階を、単回帰で1行ずつ追います。一致性に必要なのは直交条件 $\mathrm{Cov}(x, u) = 0$ であり、不偏性に必要な $\mathrm{E}(u \mid x) = 0$ までは要らないことも分かります。

3.1　ステップ1：推定量を分解する

単回帰 $y_i = \beta_1 + \beta_2 x_i + u_i$ の傾きの OLS 推定量は

$$\hat{\beta}_2 = \frac{\sum (x_i - \bar{x}) y_i}{\sum (x_i - \bar{x})^2} \tag{3.1}$$

です。y_i に真のモデルを代入して分解します。ここで、$\sum (x_i - \bar{x}) = 0$ と $\sum (x_i - \bar{x}) x_i = \sum (x_i - \bar{x})^2$ を使います。

$$\begin{aligned}
\hat{\beta}_2 &= \frac{\sum (x_i - \bar{x})(\beta_1 + \beta_2 x_i + u_i)}{\sum (x_i - \bar{x})^2} && (y_i \text{ に真のモデルを代入}) \\
&= \frac{\beta_1 \sum (x_i - \bar{x}) + \beta_2 \sum (x_i - \bar{x}) x_i + \sum (x_i - \bar{x}) u_i}{\sum (x_i - \bar{x})^2} && (\text{分子を3項に展開}) \\
&= \beta_2 + \frac{\sum (x_i - \bar{x}) u_i}{\sum (x_i - \bar{x})^2} && (\text{第1項は } 0,\ \text{第2項は } \beta_2 \textstyle\sum (x_i - \bar{x})^2) \\
&= \beta_2 + \frac{\frac{1}{N} \sum (x_i - \bar{x}) u_i}{\frac{1}{N} \sum (x_i - \bar{x})^2} && (\text{分子・分母を } N \text{ で割る})
\end{aligned}$$

最後の変形が大切です。分子・分母を N で割っても値は変わりませんが、こうすると分子・分母がどちらも「**標本平均**」の形になり、大数の法則が使えるようになります。

数式の読み方

- 第2項は「推定の誤差」です。分子は x と u の標本共分散、分母は x の標本分散です。
- 誤差の源は分子です。u_i が x_i と無相関なら、分子は 0 の周りでゆらぐだけで、データを増やすほど小さくなるはずです。

3.2 ステップ2：分子の極限

分子を $\frac{1}{N}\sum x_i u_i - \bar{x}\,\bar{u}$ と書き直します（$\sum x_i \bar{u} = N\bar{x}\bar{u}$ より）。仮定として、x_i, u_i, $x_i u_i$ の期待値の絶対値が有限であるとします。大数の法則を $x_i u_i$、x_i、u_i のそれぞれに使い、連続写像定理で積と差を処理します。

$$\frac{1}{N}\sum x_i u_i \xrightarrow{p} \mathrm{E}(xu) \qquad \text{（大数の法則）}$$

$$\bar{x} \xrightarrow{p} \mathrm{E}(x), \quad \bar{u} \xrightarrow{p} \mathrm{E}(u) \qquad \text{（大数の法則）}$$

$$\frac{1}{N}\sum x_i u_i - \bar{x}\,\bar{u} \xrightarrow{p} \mathrm{E}(xu) - \mathrm{E}(x)\mathrm{E}(u) = \mathrm{Cov}(x,u) \qquad \text{（連続写像定理：積と差）}$$

3.3 ステップ3：分母の極限と結論

分母も同様です。$\mathrm{E}(x^2) < \infty$ を仮定すると、

$$\frac{1}{N}\sum (x_i - \bar{x})^2 = \frac{1}{N}\sum x_i^2 - \bar{x}^2 \xrightarrow{p} \mathrm{E}(x^2) - (\mathrm{E}x)^2 = \mathrm{Var}(x).$$

分子が $\mathrm{Cov}(x,u)$ に、分母が $\mathrm{Var}(x)$ に確率収束するので、$\mathrm{Var}(x) > 0$（ランク条件の単回帰版）なら、$g(a,b) = a/b$ が $(\mathrm{Cov}(x,u), \mathrm{Var}(x))$ で連続であることを使って（連続写像定理）、

$$\hat{\beta}_2 \xrightarrow{p} \beta_2 + \frac{\mathrm{Cov}(x,u)}{\mathrm{Var}(x)}. \tag{3.2}$$

要点：OLS の一致性の条件（単回帰）

$\mathrm{Cov}(x,u) = 0$（直交条件）なら、$\hat{\beta}_2 \xrightarrow{p} \beta_2$、すなわち OLS は一致推定量です。$\mathrm{E}(u \mid x) = 0$ までは要りません。

数式の読み方：式 (3.2) の読み方

- $\mathrm{Cov}(x,u)/\mathrm{Var}(x)$ は「u を x に回帰したときの傾き」です。u に x と相関する部分があると、OLS はその部分を x の効果と取り違えます。
- 分母の $\mathrm{Var}(x)>0$ は、x が定数でなくばらついていることを要求します。ばらつきがなければ、傾きは識別できません。
- この取り違えの大きさは、N を増やしても小さくなりません。確率収束先がずれているからです。

計算例：直交条件が崩れたときのずれ

$\mathrm{Cov}(x,u)=0.2$、$\mathrm{Var}(x)=0.8$ なら、$\operatorname{plim}\hat{\beta}_2=\beta_2+0.2/0.8=\beta_2+0.25$ です。データを何倍に増やしても、推定値は真の値より 0.25 だけ大きいところに集まります。

3.4 内生性と欠落変数バイアス

$\mathrm{Cov}(x,u)\neq 0$ のとき、x は**内生的**であるといいます。典型例は欠落変数です。真のモデルが

$$y_i=\beta_1+\beta_2 x_i+\gamma z_i+v_i,\qquad \mathrm{Cov}(x,v)=0$$

なのに、z_i を回帰に含めなかったとします。このとき回帰の誤差は $u_i=\gamma z_i+v_i$ で、

$$\begin{aligned}\mathrm{Cov}(x,u)&=\gamma\mathrm{Cov}(x,z)+\mathrm{Cov}(x,v) && \text{（共分散の線形性）}\\ &=\gamma\mathrm{Cov}(x,z) && \text{（}\mathrm{Cov}(x,v)=0\text{）}\end{aligned}$$

となります。これを式 (3.2) に入れると

$$\operatorname{plim}\hat{\beta}_2=\beta_2+\gamma\,\frac{\mathrm{Cov}(x,z)}{\mathrm{Var}(x)}. \tag{3.3}$$

学部で学んだ欠落変数バイアスの式が、そのまま確率極限になっています。

補足

内生性があると、どれだけデータを増やしても $\hat{\beta}_2$ は β_2 に近づきません。これが、大学院で操作変数法（IV）が必要になる理由です。

3.5 行列版：一般の重回帰

説明変数が複数あるときは、x_i を定数項を含む $1\times K$ の行ベクトルとして、同じ3段階を行列で書けます。$\hat{\beta}=(\sum x_i^\top x_i)^{-1}\sum x_i^\top y_i$ に $y_i=x_i\beta+u_i$ を代入すると、

$$
\begin{aligned}
\hat{\beta} &= \beta + \left(\sum x_i^\top x_i\right)^{-1} \sum x_i^\top u_i && (y_i = x_i\beta + u_i \text{ を代入して整理}) \\
&= \beta + \left[\frac{1}{N}\sum x_i^\top x_i\right]^{-1} \frac{1}{N}\sum x_i^\top u_i && (N \text{ を2つの和から1つずつ出す}) \\
&\xrightarrow{p} \beta + \left[\mathrm{E}(x_i^\top x_i)\right]^{-1}\mathrm{E}(x_i^\top u_i) && (\text{大数の法則と連続写像定理}) \\
&= \beta && (\text{直交条件 } \mathrm{E}(x_i^\top u_i) = 0)
\end{aligned}
$$

単回帰では「平均を引く」処理が必要でしたが、x_i が定数項を含む行ベクトルなら、そのような処理は要らず、直接 $\frac{1}{N}\sum x_i^\top x_i \xrightarrow{p} \mathrm{E}(x_i^\top x_i)$ と書けます。

要点：OLS の3つの条件と役割

1. **直交条件** $\mathrm{E}(x_i^\top u_i) = 0$：一致性に使います（十分条件）。
2. $\mathrm{E}(u_i \mid x_i) = 0$：不偏性に使います。一致性だけなら不要です。
3. **ランク条件** $\mathrm{E}(x_i^\top x_i)$ が正則：逆行列に連続写像定理を使うために必要です。

さらに、大数の法則を使うための正則条件 $\mathrm{E}\|x_i^\top x_i\| < \infty$、$\mathrm{E}\|x_i^\top u_i\| < \infty$ も仮定します。これらは $\mathrm{E}|w_i| < \infty$ を各要素について満たすためのものです。

要点

- OLS 推定量は「真の値 + 標本平均の比（行列では標本平均の逆行列と標本平均の積）」に分解でき、大数の法則と連続写像定理で極限が求まります。
- 一致性には直交条件（$\mathrm{E}(x^\top u) = 0$）で十分で、不偏性に必要な $\mathrm{E}(u \mid x) = 0$ は要りません。
- ランク条件（$\mathrm{E}(x^\top x)$ が正則）は、逆行列に連続写像定理を使うために必要です。
- 直交条件が崩れる（内生性）と、$\operatorname{plim}\hat{\beta}_2 = \beta_2 + \mathrm{Cov}(x,u)/\mathrm{Var}(x)$ となり、データを増やしてもずれは消えません。

次の章へ　一致性は「推定量が真の値に近づく」ことしか言いません。第1章の2つ目の問い、「ずれ $\hat{\beta} - \beta$ はどんな分布で近似できるか」に答えるには、分布の収束を考える必要があります。

大学院とのつながり　内生性への対処が、大学院の計量経済学の大きな柱（IV・2SLS・GMM）になります。そこでも一致性の証明は、この章と同じ3段階です。

第4章

分布収束と中心極限定理

ここまでの流れ　OLS は直交条件のもとで一致推定量でした。しかし一致性は、「N を増やせば真の値に近づく」としか言いません。

この章の問い　推定量と真の値のずれは、どんな分布で近似できるのでしょうか。標本平均の「ずれ」の分布は、元のデータの分布が何であっても同じ形になるのでしょうか。

この章で分かること　なぜ $\sqrt{N}$ をかけるのか、分布収束とは何か、そして中心極限定理（CLT）です。誤差の正規性を仮定しなくても、正規分布を使った推論ができる理由が分かります。

4.1　なぜ $\sqrt{N}$ をかけるのか

大数の法則より $\bar{w}_N \xrightarrow{p} \mu$ なので、$\bar{w}_N - \mu$ はやがて 0 に潰れてしまいます（図 2.1 で分布が1本の線に縮んでいく様子です）。これでは分布の「形」が見えません。

そこで、潰れる速さに合わせて拡大します。$\mathrm{Var}(\bar{w}_N - \mu) = \sigma^2/N$ なので、$\sqrt{N}$ 倍すると

$$\begin{aligned} \mathrm{Var}\left[\sqrt{N}(\bar{w}_N - \mu)\right] &= N \cdot \mathrm{Var}(\bar{w}_N - \mu) && \text{（定数 } \sqrt{N} \text{ を二乗して外に出す）} \\ &= N \cdot \frac{\sigma^2}{N} = \sigma^2 && \text{（}\mathrm{Var}(\bar{w}_N) = \sigma^2/N\text{）} \end{aligned}$$

となり、ばらつきが 0 にも無限大にもならず、分布の形を調べられます。

数式の読み方：顕微鏡の倍率のたとえ

縮んでいく速さにちょうど合わせて顕微鏡の倍率を上げていくイメージです。倍率が低すぎる（$\sqrt{N}$ より遅い）と点に縮み、高すぎる（$\sqrt{N}$ より速い）とばらつきが無限に広がります。これが、漸近理論のいたるところに $\sqrt{N}$ が出てくる理由です。

4.2 分布収束の定義

要点：分布収束

確率変数 x の累積分布関数（CDF）を $F(\xi) = \Pr(x \le \xi)$ とします。F が連続であるすべての点 ξ で

$$\lim_{N\to\infty} \Pr(x_N \le \xi) = F(\xi) \tag{4.1}$$

が成り立つとき、x_N は x に**分布収束**するといい、$x_N \xrightarrow{d} x$ と書きます。$x \sim \mathrm{Normal}(\mu, \sigma^2)$ なら $x_N \xrightarrow{d} \mathrm{Normal}(\mu, \sigma^2)$ とも書きます。

数式の読み方

「N が大きいとき、x_N に関する確率は、x の分布を使って近似計算してよい」という意味です。例えば $x_N \xrightarrow{d} \mathrm{Normal}(0,1)$ なら、$\Pr(x_N \le 1.96) \approx \Phi(1.96) = 0.975$ と近似できます。確率収束が「値そのものが近づく」のに対して、分布収束は「**分布（確率の割り振り方）が近づく**」という、より弱い概念です。

発展：なぜ「連続である点」に限るのか

$x_N = 1/N$（確率 1 でこの値をとる）を考えます。直感的には、x_N は定数 0 に「分布収束」するべきです。極限 $x = 0$ の CDF は $F(\xi) = \mathbf{1}(\xi \ge 0)$ で、$\xi = 0$ で不連続です。ところが $\Pr(x_N \le 0) = 0$ がすべての N で成り立つので、$\xi = 0$ では $\lim \Pr(x_N \le 0) = 0 \ne 1 = F(0)$ となります。不連続点を除外しないと、この自然な例が分布収束にならなくなってしまいます。極限が正規分布のときは CDF がいたるところ連続なので、この条件を気にする必要はありません。

4.3 確率収束との関係

要点：2つの収束の関係

1. $x_N \xrightarrow{p} x$ ならば $x_N \xrightarrow{d} x$（確率収束の方が強い）。
2. 逆は一般には成り立たない。ただし、極限が**定数** c のときは同値：$x_N \xrightarrow{d} c \iff x_N \xrightarrow{p} c$。

2 の「$\Rightarrow$」を示します。定数 c の CDF は c 以外で連続なので、任意の $\varepsilon > 0$ について $\Pr(x_N \le c - \varepsilon) \to 0$、$\Pr(x_N \le c + \varepsilon/2) \to 1$ です。よって

$$\begin{aligned} \Pr(|x_N - c| > \varepsilon) &\le \Pr(x_N \le c - \varepsilon) + 1 - \Pr(x_N \le c + \varepsilon/2) && \text{（左の事象は } x_N \le c-\varepsilon \text{ または } x_N > c+\varepsilon\text{。後者は} \\ &\to 0 + 1 - 1 = 0 && \text{（上の2つの極限）} \end{aligned}$$

となります。この事実は、第5章の帰結2で使います。

4.4 中心極限定理

要点：中心極限定理（CLT）

$w_1, \ldots, w_N$ が IID、$\mathrm{E}(w_i) = \mu$、$0 < \mathrm{Var}(w_i) = \sigma^2 < \infty$ なら

$$\sqrt{N}(\bar{w}_N - \mu) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (w_i - \mu) \xrightarrow{d} \mathrm{Normal}(0, \sigma^2). \tag{4.2}$$

数式の読み方

- **驚くべき点**：w_i がどんな分布でも（左右非対称でも、離散でも）、分散さえ有限なら、標本平均の（適切に拡大した）ずれは正規分布に近づきます。だから、誤差の正規性（MLR.6）を仮定しなくても、推定量について正規分布を使った推論ができるのです。
- **同値な言い方**：両辺を σ で割ると $\dfrac{\bar{w}_N - \mu}{\sigma/\sqrt{N}} \xrightarrow{d} \mathrm{Normal}(0,1)$、つまり「標本平均を標準化すると、近似的に標準正規分布」です。
- **書き方の注意**：$\bar{w}_N$ は「近似的に $\mathrm{Normal}(\mu, \sigma^2/N)$」と言うこともありますが、これは近似の書き方です。極限の記号 $\xrightarrow{d}$ の右辺に N を含めてはいけません（極限に N は残りません）。

補足：$\mathrm{E}(w_i) = 0$ の形の CLT

CLT は $\mathrm{E}(w_i) = 0$ の形で書かれることもあります。上の定理で $w_i - \mu$ を新たに w_i と置けば同じです。

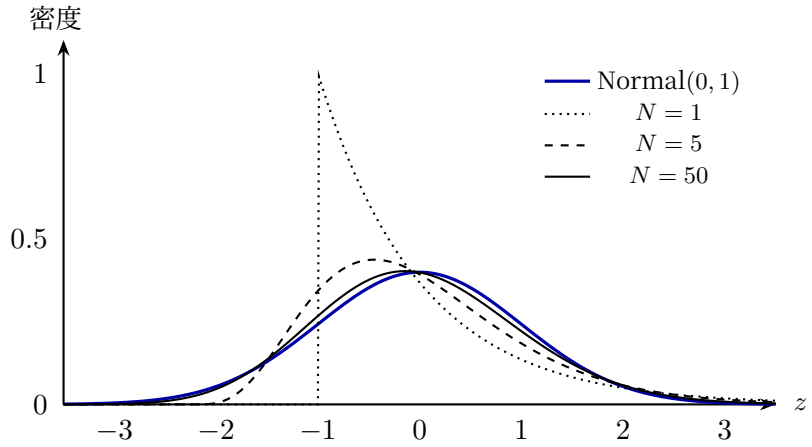
4.5 図で見る中心極限定理

指数分布（右に裾の長い非対称な分布）で、CLT の近づき方を確かめます。w_i を平均 1・分散 1 の指数分布からの標本とすると、$Z_N = \sqrt{N}(\bar{w}_N - 1)$ の密度は、ガンマ分布の密度 $g_N(s) = s^{N-1}e^{-s}/\Gamma(N)$ を使って

$$f_{Z_N}(z) = \sqrt{N}\, g_N\big(N + \sqrt{N}\, z\big) \tag{4.3}$$

と正確に書けます（$N\bar{w}_N = \sum w_i$ がガンマ分布に従うため）。

図 4.1　$\sqrt{N}(\bar{w}_N - 1)$ の密度と、CLT の極限 $\mathrm{Normal}(0,1)$



$f_{Z_N}(z) = \sqrt{N}\, g_N(N + \sqrt{N}z)$（ガンマ分布の密度）と標準正規分布の密度 $\phi(z)$ を描いた。$N = 1$ では $z > -1$ でのみ正の値をとる。

図の読み方

1. 青の太線が CLT の極限 $\mathrm{Normal}(0,1)$、黒の3本が $N = 1, 5, 50$ のときの $\sqrt{N}(\bar{w}_N - 1)$ の分布です。
2. $N = 1$（点線）は、元の指数分布の形がそのまま見えます。左端（$z = -1$）で切れ、右に長く裾をひいています。
3. $N = 5$（破線）では、左右の偏りが弱まります。$N = 50$（実線）では、青の曲線にほとんど重なります。
4. 「どれくらいの N で近似がよくなるか」は元の分布の形によって異なり、非対称性が強いほど大きな N が必要です。

4.6　ベクトル版の CLT

w_i が $K \times 1$ のベクトルのとき、分散 σ^2 は分散共分散行列 $B = \mathrm{Var}(w_i)$ に変わります（$\mathrm{E}w_i = 0$ なら $B = \mathrm{E}(w_i w_i^\top)$）。

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} w_i \xrightarrow{d} \mathrm{Normal}(0, B). \tag{4.4}$$

ここで $\mathrm{Normal}(0, B)$ は多変量正規分布です。

発展：ベクトル版 CLT と Cramér–Wold の方法

ベクトル z_N が $z \sim \text{Normal}(0, B)$ に分布収束することは、「**すべての**定数ベクトル c について $c^\top z_N \xrightarrow{d} c^\top z \sim \text{Normal}(0, c^\top B c)$」と同値です（Cramér–Wold の方法）。$c^\top w_i$ はスカラーの IID 列で分散 $c^\top B c$ なので、スカラー版 CLT から直ちにベクトル版が得られます。

4.7 連続写像定理（分布収束版）

要点：連続写像定理（分布収束版）

g が連続で $x_N \xrightarrow{d} x$ なら、$g(x_N) \xrightarrow{d} g(x)$。

計算例：二乗の例

$z_N \xrightarrow{d} \text{Normal}(0, V)$ で $g(z) = z^2/V$ とすると、$z_N^2/V \xrightarrow{d} z^2/V$ です。$z \sim \text{Normal}(0, V)$ のとき $z/\sqrt{V} \sim \text{Normal}(0, 1)$ なので、z^2/V は標準正規変数の二乗、つまり自由度 1 の χ^2 分布に従います。この形が、第7章の二次形式の最も簡単な場合です。

要点

- $\bar{w}_N - \mu$ は 0 に潰れるので、$\sqrt{N}$ をかけて拡大すると、ばらつきが一定（σ^2）になり分布の形が見えます。
- 分布収束は「確率の割り振り方が近づく」こと、確率収束は「値そのものが近づく」ことで、確率収束 $\Rightarrow$ 分布収束です。極限が定数のときは両者は同値です。
- CLT：$\sqrt{N}(\bar{w}_N - \mu) \xrightarrow{d} \text{Normal}(0, \sigma^2)$。元の分布が正規でなくても、分散が有限なら成り立ちます。
- 分布収束版の連続写像定理により、$g(x_N) \xrightarrow{d} g(x)$（$g$ は連続）。

次の章へ　CLT は標本平均についての定理です。OLS 推定量は、標本平均の比（分子は CLT、分母は大数の法則）の形をしています。この2つをつなぐのが、次の章の Slutsky の定理です。

大学院とのつながり　大学院の講義の漸近正規性の主張は、すべて $\sqrt{N}(\hat{\beta} - \beta) \xrightarrow{d} \text{Normal}(0, \cdot)$ の形で書かれます。

第5章

Slutsky の定理と OLS の漸近正規性

ここまでの流れ　標本平均については、大数の法則（確率収束）と中心極限定理（分布収束）がそろいました。
この章の問い　OLS 推定量のずれ $\sqrt{N}(\hat{\beta}-\beta)$ は、「CLT で分布収束する部分」と「大数の法則で確率収束する部分」の比になっています。この2つはどう組み合わせればよいのでしょうか。
この章で分かること　分布収束と確率収束を組み合わせる Slutsky の定理、その2つの帰結、そして OLS の漸近正規性（サンドイッチ型の漸近分散）の証明です。

5.1　Slutsky の定理

要点：Slutsky の定理（多くの教科書での意味）

$x_N \xrightarrow{d} x$、$y_N \xrightarrow{p} c$（**定数**）なら

$$x_N + y_N \xrightarrow{d} x + c, \qquad y_N x_N \xrightarrow{d} cx, \qquad x_N / y_N \xrightarrow{d} x/c \ \ (c \neq 0). \tag{5.1}$$

ベクトル・行列でも同様で、$x_N \xrightarrow{d} x$（$K \times 1$）、$Y_N \xrightarrow{p} C$（定数行列）なら $Y_N x_N \xrightarrow{d} Cx$ です。

数式の読み方：直感

y_N は N が大きければほぼ定数 c に等しいので、「ほぼ定数」を「定数」に置き換えても、分布の極限は変わらない、ということです。

補足：つまずきポイント

y_N の極限が**定数**であることが本質的です。$x_N \xrightarrow{d} x$、$y_N \xrightarrow{d} y$（両方とも確率変数に分布収束）だけでは、$x_N + y_N \xrightarrow{d} x + y$ とは一般に言えません。分布収束は「それぞれの分布」の情報しか持たず、x_N と y_N の同時の振る舞い（相関など）が分からないからです。

計算例：標本平均の t 統計量

w_i が IID で平均 μ、分散 $\sigma^2 > 0$ とします。$s_N^2 = \frac{1}{N-1}\sum(w_i - \bar{w}_N)^2 \xrightarrow{p} \sigma^2$ ですから、連続写像定理で $s_N \xrightarrow{p} \sigma$ です。$\sqrt{N}(\bar{w}_N - \mu) \xrightarrow{d} \text{Normal}(0, \sigma^2)$ に Slutsky の定理（割り算）を使うと、

$$\frac{\sqrt{N}(\bar{w}_N - \mu)}{s_N} \xrightarrow{d} \frac{\text{Normal}(0, \sigma^2)}{\sigma} = \text{Normal}(0, 1)$$

となります。分母の σ が未知で推定値に置き換えても、極限の分布は変わらないということです。

5.2 特に重要な2つの帰結

要点：Slutsky の定理の2つの帰結

1. **0 に確率収束する余分な項は、分布の極限に影響しない**：$x_N = z_N + r_N$、$z_N \xrightarrow{d} z$、$r_N \xrightarrow{p} 0$ なら $x_N \xrightarrow{d} z$。
2. **0 に確率収束するもの × 分布収束するもの → 0 に確率収束**：$y_N \xrightarrow{p} 0$、$x_N \xrightarrow{d} x$ なら $y_N x_N \xrightarrow{d} 0 \cdot x = 0$。極限が定数なので、第4章の結果より $y_N x_N \xrightarrow{p} 0$ です。

帰結1は定理の和の部分で $c = 0$ としたもの、帰結2は積の部分で $c = 0$ としたものです。第9章で扱う o_p, O_p の言葉では、帰結2は「$o_p(1) \cdot O_p(1) = o_p(1)$」にあたります。

補足：帰結1の書き方

「$x_N - z_N \xrightarrow{p} 0$ かつ $z_N \xrightarrow{d} z$ なら $x_N \xrightarrow{d} z$」という形で述べられることもあります。これは帰結1（$x_N = z_N + (x_N - z_N)$ と書く）です。

5.3 OLS の漸近正規性（単回帰）

第3章で一致性を示したのと同じ単回帰で、$\sqrt{N}(\hat{\beta}_2 - \beta_2)$ の分布を求めます。

要点：仮定

(x_i, u_i) は IID、$\text{E}(u) = 0$、$\text{E}(xu) = 0$（直交条件）、$0 < \text{Var}(x) < \infty$、$\text{Var}(u) < \infty$。さらに、$\mu_x = \text{E}(x)$ として、$B = \text{E}[(x - \mu_x)^2 u^2]$ が有限で正であるとします。

ステップ1：$\sqrt{N}$ をかけた分解　第3章のステップ1の式に $\sqrt{N}$ をかけます。分子の $\frac{1}{N}$ と $\sqrt{N}$ が合わさって $\frac{1}{\sqrt{N}}$ になります。

$$\sqrt{N}(\hat{\beta}_2 - \beta_2) = \frac{\dfrac{1}{\sqrt{N}}\sum(x_i - \bar{x})u_i}{\dfrac{1}{N}\sum(x_i - \bar{x})^2}. \tag{5.2}$$

ステップ2：分子を「CLT が使える部分」と「余分な部分」に分ける　$x_i - \bar{x} = (x_i - \mu_x) - (\bar{x} - \mu_x)$ と書き直すと、分子は次のように分かれます。

$$\begin{aligned}
\frac{1}{\sqrt{N}}\sum(x_i - \bar{x})u_i &= \frac{1}{\sqrt{N}}\sum\big[(x_i - \mu_x) - (\bar{x} - \mu_x)\big]u_i && (x_i - \bar{x} \text{ を書き換える}) \\
&= \underbrace{\frac{1}{\sqrt{N}}\sum(x_i - \mu_x)u_i}_{(a)} - \underbrace{(\bar{x} - \mu_x)}_{\xrightarrow{p}0} \cdot \underbrace{\frac{1}{\sqrt{N}}\sum u_i}_{\xrightarrow{d}\text{Normal}(0,\text{Var}u)} && ((\bar{x} - \mu_x) \text{ は } i \text{ によらないので和の外へ})
\end{aligned}$$

数式の読み方：なぜ分けるのか

$\bar{x}$ はデータ全体から計算されるので、$(x_i - \bar{x})u_i$ は i について独立ではなく、CLT を直接使えません。真の平均 μ_x で置き換えた (a) なら、IID の和になります。余った部分は 0 に潰れる量と、ふつうの大きさの量の積になっています。

ステップ3：各部分の極限

- **(a)**：$w_i = (x_i - \mu_x)u_i$ は IID で、平均は $\mathrm{E}(xu) - \mu_x\mathrm{E}(u) = 0$、分散は $\mathrm{E}(w_i^2) = B$ です。CLT より (a) $\xrightarrow{d}$ Normal$(0, B)$。
- **余分な項**：大数の法則より $\bar{x} - \mu_x \xrightarrow{p} 0$、CLT より $\frac{1}{\sqrt{N}}\sum u_i \xrightarrow{d}$ Normal$(0, \text{Var}u)$。帰結2より、積は $\xrightarrow{p} 0$。
- **分母**：第3章と同じく $\frac{1}{N}\sum(x_i - \bar{x})^2 \xrightarrow{p} \text{Var}(x)$。

ステップ4：組み立てる　帰結1より、分子は $\xrightarrow{d}$ Normal$(0, B)$ です。分母は $\text{Var}(x) > 0$ に確率収束するので、Slutsky の定理（割り算）より

$$\begin{aligned}
\sqrt{N}(\hat{\beta}_2 - \beta_2) &\xrightarrow{d} \frac{1}{\text{Var}(x)}\,\text{Normal}(0, B) && (\text{Slutsky の定理（割り算）}) \\
&= \text{Normal}\left(0,\ \frac{B}{[\text{Var}(x)]^2}\right) && (\text{定数 } a \text{ 倍すると分散は } a^2 \text{ 倍})
\end{aligned}$$

すなわち

$$\sqrt{N}(\hat{\beta}_2 - \beta_2) \xrightarrow{d} \text{Normal}\left(0,\ \frac{\mathrm{E}[(x - \mu_x)^2 u^2]}{[\text{Var}(x)]^2}\right). \tag{5.3}$$

数式の読み方：サンドイッチ型

漸近分散 $B/A^2 = A^{-1}BA^{-1}$（$A = \mathrm{Var}(x)$、$B = \mathrm{E}[(x-\mu_x)^2u^2]$）は、$B$ を A^{-1} で両側から挟んだ形なので、**サンドイッチ型**と呼ばれます。A は x のばらつき（分母）、B は「x の偏差で重みづけた誤差のばらつき」（分子の分散）です。

5.4 均一分散の場合

誤差の条件付き分散が x によらない、すなわち $\mathrm{E}(u^2 \mid x) = \sigma^2$ のとき、繰り返し期待値の法則で B を簡単にできます。

$$
\begin{aligned}
B &= \mathrm{E}\big[(x-\mu_x)^2u^2\big] \\
&= \mathrm{E}\Big[(x-\mu_x)^2\,\mathrm{E}(u^2 \mid x)\Big] && \text{（繰り返し期待値の法則）} \\
&= \sigma^2\mathrm{E}\big[(x-\mu_x)^2\big] = \sigma^2\mathrm{Var}(x) && \text{（}\mathrm{E}(u^2 \mid x)=\sigma^2\text{ を代入）}
\end{aligned}
$$

したがって、漸近分散は $B/[\mathrm{Var}(x)]^2 = \sigma^2/\mathrm{Var}(x)$ に簡単になります。N で割ると $\sigma^2/(N\mathrm{Var}(x)) \approx \sigma^2/\sum(x_i-\bar{x})^2$ で、学部で学んだ分散の公式と一致します。

補足

均一分散が成り立たないとき、$B \neq \sigma^2\mathrm{Var}(x)$ となり、学部の分散の公式は使えません。サンドイッチ型 B/A^2 はそのまま成り立ちます。これが、頑健な標準誤差（第6章）が必要な理由です。

5.5 行列版

一般の重回帰（x_i は定数項を含む $1 \times K$ の行ベクトル）では、第3章の分解に $\sqrt{N}$ をかけて

$$
\sqrt{N}(\hat{\beta}-\beta) = \Big[\frac{1}{N}\sum x_i^\top x_i\Big]^{-1}\frac{1}{\sqrt{N}}\sum x_i^\top u_i \tag{5.4}
$$

とします。第1の因子は大数の法則と連続写像定理で A^{-1}（$A = \mathrm{E}(x_i^\top x_i)$）に確率収束し、第2の因子はベクトル版 CLT（$\mathrm{E}(x_i^\top u_i) = 0$ を使う）で $\mathrm{Normal}(0, B)$（$B = \mathrm{E}(u_i^2 x_i^\top x_i)$）に分布収束します。Slutsky の定理（行列版）より

$$
\sqrt{N}(\hat{\beta}-\beta) \xrightarrow{d} A^{-1}\mathrm{Normal}(0,B) = \mathrm{Normal}(0,\ A^{-1}BA^{-1}). \tag{5.5}
$$

最後の等号は、正規分布に従うベクトルに行列 A^{-1} をかけると、分散が $A^{-1}B(A^{-1})^\top$ になることによります（A は対称なので $(A^{-1})^\top = A^{-1}$。第7章）。

要点

- Slutsky の定理：$x_N \xrightarrow{d} x$ と $y_N \xrightarrow{p} c$（定数）なら、和・積・商の極限は x と c で計算できます。y_N が定数に収束することが本質的です。
- 帰結1：0 に確率収束する余分な項は無視できます。帰結2：0 に確率収束するものに分布収束するものをかけると 0 に確率収束します。
- OLS の漸近正規性の証明：分解 → 分子に CLT・分母に大数の法則 → Slutsky の定理の組み立て、の3段階です。
- 漸近分散はサンドイッチ型 $A^{-1}BA^{-1}$。均一分散のとき $\sigma^2 A^{-1}$（単回帰では $\sigma^2/\mathrm{Var}(x)$）に簡単になります。

次の章へ　漸近分散 $A^{-1}BA^{-1}$ は未知の量なので、データから推定して標準誤差を作る必要があります。次の章では、推定した分散で作った t 統計量がなぜ標準正規分布に従うのかを示します。

大学院とのつながり　$A^{-1}BA^{-1}$ の形は、GLS・IV・GMM でも形を変えて何度も登場します。

第6章

漸近分散・標準誤差・t 統計量

ここまでの流れ　OLS 推定量は $\sqrt{N}(\hat{\beta}-\beta) \xrightarrow{d} \text{Normal}(0, A^{-1}BA^{-1})$ を満たすことが分かりました。

この章の問い　漸近分散 $A^{-1}BA^{-1}$ は未知の量です。これを推定して作った標準誤差で t 統計量を作ると、なぜ「$|t| > 1.96$ なら棄却」が使えるのでしょうか。

この章で分かること　漸近分散と標準誤差の定義、t 統計量が標準正規分布に近づく理由（Slutsky の定理）、そして頑健な標準誤差が必要な理由です。

6.1　漸近分散と標準誤差

要点：漸近分散と標準誤差

$\sqrt{N}(\hat{\theta}_N - \theta) \xrightarrow{d} \text{Normal}(0, V)$ のとき、V を**漸近分散**といい、

$$\text{Avar}\left[\sqrt{N}(\hat{\theta}_N - \theta)\right] = V, \qquad \text{または} \qquad \text{Avar}(\hat{\theta}_N) = V/N$$

と書きます。V の一致推定量 $\hat{V}$ を用いて、$\sqrt{\hat{V}/N}$（ベクトルなら $\hat{V}/N$ の対角要素の平方根）を**標準誤差**といい、$\text{se}(\hat{\theta}_N)$ と書きます。

補足：つまずきポイント

$\text{Avar}(\hat{\theta}_N) = V/N$ は、「N が大きいとき $\hat{\theta}_N$ の分布は、平均 θ、分散 V/N の正規分布で近似できる」という近似の表記です。$\hat{\theta}_N$ の本当の分散 $\text{Var}(\hat{\theta}_N)$ が V/N に等しいという意味ではありません（本当の分散は存在しないことさえあります）。

6.2　t 統計量が標準正規分布に近づく理由

$V > 0$ とします。t 統計量を、$\sqrt{N}$ を使って書き直します。

$$\begin{aligned} t &= \frac{\hat{\theta}_N - \theta}{\text{se}(\hat{\theta}_N)} && (t \text{ 統計量の定義}) \\ &= \frac{\hat{\theta}_N - \theta}{\sqrt{\hat{V}/N}} && (\text{se}(\hat{\theta}_N) = \sqrt{\hat{V}/N}) \\ &= \frac{\sqrt{N}(\hat{\theta}_N - \theta)}{\sqrt{\hat{V}}} && (\text{分子・分母に } \sqrt{N} \text{ をかける}) \end{aligned}$$

分子と分母を別々に見ます。

- **分子**：$\sqrt{N}(\hat{\theta}_N - \theta) \xrightarrow{d} \mathrm{Normal}(0, V)$（漸近正規性）。
- **分母**：$\hat{V} \xrightarrow{p} V > 0$ で、$\sqrt{\cdot}$ は $V > 0$ で連続なので、連続写像定理より $\sqrt{\hat{V}} \xrightarrow{p} \sqrt{V}$。

Slutsky の定理（割り算）より

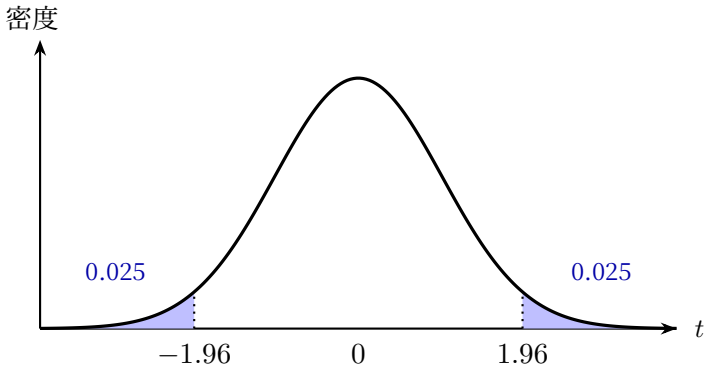
$$t \xrightarrow{d} \frac{1}{\sqrt{V}}\mathrm{Normal}(0, V) = \mathrm{Normal}(0, 1). \tag{6.1}$$

数式の読み方

- t 統計量は「推定値が仮説の値から標準誤差の何個分離れているか」です。
- 分子の分布は未知の V に依存しますが、分母で $\sqrt{\hat{V}}$ を割ると、V の大きさが打ち消され、$\mathrm{Normal}(0, 1)$ という分布が残ります。
- 標準誤差に使う $\hat{V}$ が V **の一致推定量**であることが、本質的に重要です。

これが、帰無仮説 $H_0 : \theta = \theta_0$ のもとで「$|t| > 1.96$ なら棄却」という検定が使える理由です。

図 6.1　標準正規分布と、$|t| > 1.96$ の棄却域



曲線は標準正規分布の密度 $\phi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$。$\phi(1.96) \approx 0.0584$。両側の裾の面積はそれぞれ約 0.025。

図の読み方

1. 黒の曲線は、帰無仮説が正しいときの t 統計量の極限の分布 $\mathrm{Normal}(0,1)$ です。
2. 青く塗った部分が棄却域（$|t| > 1.96$）です。左右の裾の面積は、どちらも約 0.025 です。
3. 帰無仮説が正しいのに棄却してしまう確率（漸近サイズ）は、合計で約 0.05 です。

計算例：検定の手順

$\hat{\theta}_N = 0.52$、$\mathrm{se}(\hat{\theta}_N) = 0.2$ で、$H_0 : \theta = 0$ を検定するとします。$t = 0.52/0.2 = 2.6$ で、$|t| = 2.6 > 1.96$ なので、5% の漸近的な有意水準で H_0 を棄却します。

6.3　頑健な標準誤差が必要な理由

均一分散を仮定して作った $\hat{V}_h$ は、不均一分散のもとでは、真の漸近分散 V に収束せず、別の値 V_h に収束します（第5章：サンドイッチの B が $\sigma^2\mathrm{Var}(x)$ にならないため）。このとき、上と同じ計算で

$$t = \frac{\sqrt{N}(\hat{\theta}_N - \theta)}{\sqrt{\hat{V}_h}} \xrightarrow{d} \frac{\mathrm{Normal}(0, V)}{\sqrt{V_h}} = \mathrm{Normal}\left(0,\ \frac{V}{V_h}\right) \tag{6.2}$$

となり、$V \neq V_h$ なら t は標準正規分布に近づきません。「$|t| > 1.96$ で棄却」を使っても、棄却の確率が 5% になりません。

要点：頑健な標準誤差

サンドイッチ型の漸近分散 $A^{-1}BA^{-1}$ の A, B を、それぞれ標本平均 $\hat{A}, \hat{B}$ で置き換えた $\hat{V} = \hat{A}^{-1}\hat{B}\hat{A}^{-1}$ から作った標準誤差を**頑健な標準誤差**といいます。均一分散を仮定せずに V の一致推定量になるので、t 統計量が $\text{Normal}(0,1)$ に近づきます。

発展：頑健な分散推定量の一致性

単回帰では $\hat{B} = \frac{1}{N}\sum(x_i - \bar{x})^2\hat{u}_i^2$、$\hat{A} = \frac{1}{N}\sum(x_i - \bar{x})^2$ として $\hat{V} = \hat{B}/\hat{A}^2$ です。残差は $\hat{u}_i = u_i - (\hat{\beta}_1 - \beta_1) - (\hat{\beta}_2 - \beta_2)x_i$ と書けます。これを代入して展開すると、$\hat{B}$ は「$\frac{1}{N}\sum(x_i - \mu_x)^2u_i^2$（大数の法則で B に収束）＋（$\hat{\beta} - \beta = o_p(1)$ がかかった項）」に分かれ、後者は適当なモーメント条件（x と u の4次のモーメントが有限など）のもとで $o_p(1)$ になります（o_p の記法は第9章）。

要点

- 漸近分散 V は $\sqrt{N}(\hat{\theta} - \theta)$ の極限分布の分散です。$\text{Avar}(\hat{\theta}) = V/N$ は近似の表記で、本当の分散ではありません。
- 標準誤差は $\sqrt{\hat{V}/N}$（$\hat{V}$ は V の一致推定量）です。
- $t = \sqrt{N}(\hat{\theta} - \theta)/\sqrt{\hat{V}}$ は、分子が $\text{Normal}(0, V)$ に分布収束し、分母が $\sqrt{V}$ に確率収束するので、Slutsky の定理で $\text{Normal}(0,1)$ に近づきます。
- 均一分散を仮定した $\hat{V}$ は、不均一分散のもとでは V に収束しないため、頑健な標準誤差を使います。

次の章へ　t 検定は制約が1本の検定です。複数の制約を同時に検定する（例：2つの係数が両方 0）には、ベクトルの分布収束と二次形式を使います。

大学院とのつながり　大学院の講義では、頑健な標準誤差を標準として使います。$\hat{V}$ の一致性を確かめることが、検定の正当化の中心になります。

第7章

ベクトルの場合：線形変換・二次形式・Wald 検定

ここまでの流れ　係数が1つの場合（t 検定）の漸近理論はそろいました。推定量がベクトル $\hat{\theta}$ のとき、$\sqrt{N}(\hat{\theta}-\theta) \xrightarrow{d} \text{Normal}(0, V)$ です。
この章の問い　複数の制約（例：「2つの係数が両方 0」）をまとめて検定するには、どんな統計量を作り、どの分布で判断すればよいのでしょうか。
この章で分かること　多変量正規分布の2つの性質、線形変換と二次形式の分布収束、そして Wald 統計量が χ^2 分布に近づく理由です。

7.1　多変量正規分布の2つの性質

以下では、多変量正規分布の標準的な性質を使います。$z \sim \text{Normal}(0, V)$（$K \times 1$、$V$ は正定値）とします。

要点：多変量正規分布の2つの性質

1. **線形変換しても正規**：定数行列 R（$Q \times K$）について、$Rz \sim \text{Normal}(0, RVR^\top)$。分散は公式 $\text{Var}(Aw) = A\,\text{Var}(w)A^\top$ から分かります。
2. **無相関なら独立**：多変量正規分布の要素どうしは、共分散が 0 なら独立です。一般の分布では「無相関 $\Rightarrow$ 独立」は成り立たないので、これは正規分布の特別な性質です。

7.2　線形変換

z_N が $K \times 1$ で $z_N \xrightarrow{d} \text{Normal}(0, V)$、$V$ は正定値とします。定数行列 R（$Q \times K$）について

$$Rz_N \xrightarrow{d} \text{Normal}(0,\, RVR^\top). \tag{7.1}$$

理由は、$g(z) = Rz$ が連続なので、分布収束版の連続写像定理より $Rz_N \xrightarrow{d} Rz$ であり、性質1より $Rz \sim \text{Normal}(0, RVR^\top)$ だからです。

第5章で OLS の漸近分散を求めるときに使った変形（$A^{-1}\text{Normal}(0, B) = \text{Normal}(0, A^{-1}BA^{-1})$）は、この線形変換の特別な場合です。

7.3 二次形式

要点：二次形式

$z_N \xrightarrow{d} \text{Normal}(0, V)$（$K \times 1$、$V$ は正定値）なら

$$z_N^\top V^{-1} z_N \xrightarrow{d} \chi^2_K. \tag{7.2}$$

さらに、$\hat{V} \xrightarrow{p} V$ なら、推定した $\hat{V}$ を使っても同じです。

$$z_N^\top \hat{V}^{-1} z_N \xrightarrow{d} \chi^2_K. \tag{7.3}$$

V を使った二次形式の理由　分布収束版の連続写像定理（$g(z) = z^\top V^{-1} z$ は連続）より、$z_N^\top V^{-1} z_N \xrightarrow{d} z^\top V^{-1} z$（$z \sim \text{Normal}(0, V)$）です。$V$ の（対称な）平方根 $V^{1/2}$（$V^{1/2} V^{1/2} = V$）とその逆行列 $V^{-1/2}$ を使って $e = V^{-1/2} z$ とおきます。

$$\begin{aligned} e &\sim \text{Normal}\left(0,\ V^{-1/2} V V^{-1/2}\right) && \text{（性質1：線形変換）} \\ &= \text{Normal}(0,\ I_K) && (V^{-1/2} V V^{-1/2} = I_K) \end{aligned}$$

e の要素は分散 1 で互いに無相関で、性質2より互いに独立な標準正規変数です。したがって

$$\begin{aligned} z^\top V^{-1} z &= (V^{-1/2} z)^\top (V^{-1/2} z) && (V^{-1} = V^{-1/2} V^{-1/2}) \\ &= e^\top e = e_1^2 + \cdots + e_K^2 \sim \chi^2_K && \text{（独立な標準正規変数の二乗和は } \chi^2_K \text{ の定義）} \end{aligned}$$

数式の読み方

- $z^\top V^{-1} z$ は、z の「大きさ」を、各方向のばらつき V で割って測った量です。$K = 1$ なら z^2/V、つまり標準化した値の二乗です（第4章の計算例）。
- V^{-1} で割るのは、ばらつきの大きい方向の大きな値と、相関のある方向の同じ情報の二重カウントを、調整するためです。

V を $\hat{V}$ に置き換えてよい理由

$$z_N^\top \hat{V}^{-1} z_N - z_N^\top V^{-1} z_N = z_N^\top \left(\hat{V}^{-1} - V^{-1}\right) z_N \qquad \text{（共通の左右の因子 } z_N \text{ をくくり出す）}$$

$\hat{V} \xrightarrow{p} V$ で V は正則なので、連続写像定理より $\hat{V}^{-1} - V^{-1} \xrightarrow{p} 0$ です。z_N は分布収束するので、Slutsky の定理の帰結2より、上の差全体が $\xrightarrow{p} 0$ です。帰結1より、$z_N^\top \hat{V}^{-1} z_N$ の極

限の分布は $z_N^\top V^{-1} z_N$ と同じです。

7.4 Wald 検定

推定量 $\hat{\theta}$（$K \times 1$）が $\sqrt{N}(\hat{\theta} - \theta) \xrightarrow{d} \text{Normal}(0, V)$ を満たすとします。Q 本の線形制約

$$H_0 : \ R\theta = r \qquad (R \text{ は } Q \times K,\ \text{rank}\, R = Q)$$

を検定します。H_0 のもとでは $R\theta = r$ なので、

$$\begin{aligned}
\sqrt{N}(R\hat{\theta} - r) &= \sqrt{N}(R\hat{\theta} - R\theta) && (H_0 \text{ のもとで } r = R\theta) \\
&= R\sqrt{N}(\hat{\theta} - \theta) && (R \text{ をくくり出す}) \\
&\xrightarrow{d} \text{Normal}(0,\ RVR^\top) && (\text{線形変換})
\end{aligned}$$

です。$\text{rank}\, R = Q$ と V の正定値性から、$RVR^\top$（$Q \times Q$）は正定値です。そこで、前の節の二次形式を $z_N = \sqrt{N}(R\hat{\theta} - r)$、分散 $RVR^\top$ に適用します。

要点：Wald 統計量

$$W = (R\hat{\theta} - r)^\top \left[R(\hat{V}/N)R^\top\right]^{-1} (R\hat{\theta} - r) \xrightarrow{d} \chi^2_Q \qquad (H_0 \text{ のもと}). \tag{7.4}$$

最初の等号が成り立つ理由は、$[R(\hat{V}/N)R^\top]^{-1} = N(R\hat{V}R^\top)^{-1}$ なので、$N = \sqrt{N} \cdot \sqrt{N}$ を両側のベクトルに分配できるからです。

$$\begin{aligned}
W &= (R\hat{\theta} - r)^\top N \, (R\hat{V}R^\top)^{-1} (R\hat{\theta} - r) && ([R(\hat{V}/N)R^\top]^{-1} = N(R\hat{V}R^\top)^{-1}) \\
&= \left[\sqrt{N}(R\hat{\theta} - r)\right]^\top (R\hat{V}R^\top)^{-1} \left[\sqrt{N}(R\hat{\theta} - r)\right] && (N = \sqrt{N} \cdot \sqrt{N} \text{ を両側に分配}) \\
&\xrightarrow{d} \chi^2_Q && (\text{二次形式（}\hat{V} \text{ を使った版）})
\end{aligned}$$

数式の読み方

- $R\hat{\theta} - r$ は、制約が正しければ 0 に近いはずのベクトルです。その大きさを、ばらつき $R(\hat{V}/N)R^\top$ で測ったものが W です。
- 自由度 Q は制約の本数です。「2つの係数が両方 0」なら $R = I_2$、$r = 0$、$Q = 2$ です。
- 標準的な漸近的検定は、W が χ^2_Q の上側 5% 点を超えたら棄却します。統計表の値は、$Q = 1$ で約 3.84（$= 1.96^2$）、$Q = 2$ で約 5.99、$Q = 3$ で約 7.81 です。

補足：制約が1本のとき

$Q = 1$ なら、Wald 統計量は t 統計量の二乗に一致します。したがって、5% 水準の Wald 検定と両側 t 検定は同じ結論になります（練習問題で確かめます）。

要点

- $z_N \xrightarrow{d} \text{Normal}(0, V)$ のとき、線形変換は $Rz_N \xrightarrow{d} \text{Normal}(0, RVR^\top)$ です。
- 二次形式 $z_N^\top V^{-1} z_N \xrightarrow{d} \chi^2_K$。標準化して無相関・独立な標準正規にそろえるのがポイントです。
- V を一致推定量 $\hat{V}$ に置き換えても極限は変わりません（連続写像定理と Slutsky の定理の帰結）。
- Wald 統計量 W は、$H_0 : R\theta = r$ のもとで χ^2_Q に分布収束します。

次の章へ　ここまでは、推定量そのものの分布を扱いました。次の章では、推定量の**関数**（例：2つの係数の比、弾力性）の分布を、デルタ法で求めます。

大学院とのつながり　Wald 検定は、大学院の講義で係数の複数制約の検定に標準的に使います。GMM の過剰識別制約の検定も、同じ二次形式の発想です。

第8章

デルタ法

ここまでの流れ　推定量 $\hat{\theta}$ の漸近正規性 $\sqrt{N}(\hat{\theta}-\theta) \xrightarrow{d} \mathrm{Normal}(0, V)$ が分かり、t 検定や Wald 検定ができるようになりました。

この章の問い　関心のある量が、パラメータそのものではなく、その関数（例：θ^2、比 θ_1/θ_2、弾力性）のときは、推定量 $c(\hat{\theta})$ はどんな分布に近づくのでしょうか。

この章で分かること　デルタ法の主張と、平均値の定理を使った証明、接線による直感、そしてベクトル版（ヤコビ行列）です。

8.1　デルタ法の主張

要点：デルタ法（スカラー版）

$\sqrt{N}(\hat{\theta}-\theta) \xrightarrow{d} \mathrm{Normal}(0, V)$ で、関数 c が θ の近くで連続微分可能なら

$$\sqrt{N}\big[c(\hat{\theta}) - c(\theta)\big] \xrightarrow{d} \mathrm{Normal}\big(0,\ c'(\theta)^2 V\big). \tag{8.1}$$

数式の読み方

- $c'(\theta)$ は、θ が少し動いたとき $c(\theta)$ がどれだけ動くかを表す「拡大率」です。
- $\hat{\theta}$ のばらつき（標準偏差）が $\sqrt{V}$ なら、$c(\hat{\theta})$ のばらつきは $|c'(\theta)|\sqrt{V}$、つまり分散は $c'(\theta)^2 V$ 倍されます。

計算例：2つの例

(1) 二乗：$\sqrt{N}(\bar{w}-\mu) \xrightarrow{d} \mathrm{Normal}(0, \sigma^2)$ のとき、$c(\mu)=\mu^2$ なら $c'(\mu)=2\mu$ です。したがって

$$\sqrt{N}(\bar{w}^2 - \mu^2) \xrightarrow{d} \mathrm{Normal}(0,\ 4\mu^2\sigma^2).$$

(2) 対数：$c(\theta)=\log\theta$（$\theta>0$）なら $c'(\theta)=1/\theta$ で、$\sqrt{N}(\log\hat{\theta}-\log\theta) \xrightarrow{d} \mathrm{Normal}(0, V/\theta^2)$ です。対数の差は変化率に近いので、「標準誤差を θ で割った相対的なばらつき」が、対数のばらつきになるということです。

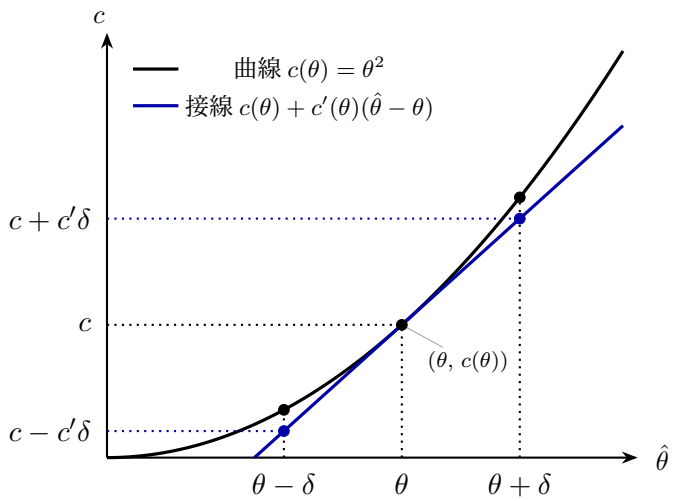
8.2 直感：接線で近似する

N が大きいと、$\hat{\theta}$ は θ のすぐ近くにあります。その狭い範囲では、c はほぼ直線（θ での接線）で近似できます。

$$c(\hat{\theta}) \approx c(\theta) + c'(\theta)(\hat{\theta} - \theta). \tag{8.2}$$

これはテイラー展開の1次近似です。直線による変換なら、正規分布はそれを傾き倍に拡大するだけです。これが式 (8.1) の中身です。

図 8.1　曲線 $c(\theta) = \theta^2$ と、$\theta = 1$ での接線



$c(\theta) = \theta^2$、$\theta = 1$ での接線は $2\hat{\theta} - 1$（$c'(1) = 2$）。$\delta = 0.4$ とした。曲線上の点は (0.6, 0.36)、(1.4, 1.96)、接線上の点は (0.6, 0.2)、(1.4, 1.8)。

図の読み方

1. 黒い曲線が $c(\theta) = \theta^2$、青い直線が $\theta = 1$ での接線です。横軸が推定量 $\hat{\theta}$ の値、縦軸が $c(\hat{\theta})$ の値です。

2. $\hat{\theta}$ が $\theta \pm \delta$ の範囲にあるとき、接線で見た c の範囲は $c(\theta) \pm c'(\theta)\delta$（青の点線）です。範囲の幅が $|c'(\theta)|$ 倍になるので、ばらつきも $|c'(\theta)|$ 倍、分散は $c'(\theta)^2$ 倍です。

3. δ が大きい（θ から離れる）と、曲線と接線はずれます（図の黒の点と青の点）。N が大きいと $\hat{\theta}$ は θ のごく近くにしか現れないので、このずれは無視できるようになります。

8.3 証明

接線による近似を、平均値の定理で厳密にします。

ステップ1：平均値の定理　平均値の定理より、$\hat{\theta}$ と θ の間のある点 $\bar{\theta}$ について

$$\begin{aligned} c(\hat{\theta}) - c(\theta) &= c'(\bar{\theta})(\hat{\theta} - \theta) && \text{（平均値の定理）} \\ \sqrt{N}\big[c(\hat{\theta}) - c(\theta)\big] &= c'(\bar{\theta}) \cdot \sqrt{N}(\hat{\theta} - \theta) && \text{（両辺に } \sqrt{N} \text{ をかける）} \end{aligned}$$

ステップ2：$\bar{\theta} \xrightarrow{p} \theta$ を示す　まず $\hat{\theta} \xrightarrow{p} \theta$ を示します（漸近正規性から一致性が従います）。

$$\hat{\theta} - \theta = \frac{1}{\sqrt{N}} \cdot \sqrt{N}(\hat{\theta} - \theta)$$

で、$\frac{1}{\sqrt{N}} \to 0$（定数列）、$\sqrt{N}(\hat{\theta} - \theta)$ は分布収束するので、第5章の帰結2より $\hat{\theta} - \theta \xrightarrow{p} 0$ です。$\bar{\theta}$ は $\hat{\theta}$ と θ の間にあるので $|\bar{\theta} - \theta| \le |\hat{\theta} - \theta|$、よって $\bar{\theta} \xrightarrow{p} \theta$ です。

ステップ3：組み立てる　c' が連続なので、連続写像定理より $c'(\bar{\theta}) \xrightarrow{p} c'(\theta)$ です。$\sqrt{N}(\hat{\theta} - \theta) \xrightarrow{d} \mathrm{Normal}(0, V)$ なので、Slutsky の定理（掛け算）より

$$\sqrt{N}\big[c(\hat{\theta}) - c(\theta)\big] = c'(\bar{\theta}) \cdot \sqrt{N}(\hat{\theta} - \theta) \xrightarrow{d} c'(\theta) \cdot \mathrm{Normal}(0, V) = \mathrm{Normal}\big(0,\ c'(\theta)^2 V\big).$$

8.4 ベクトル版

要点：デルタ法（ベクトル版）

$c : \mathbb{R}^K \to \mathbb{R}^Q$ のとき、$c'(\theta)$ を $Q \times K$ の**ヤコビ行列**

$$C(\theta) = \frac{\partial c(\theta)}{\partial \theta^\top} \qquad \big((q,k) \text{ 要素は } \partial c_q / \partial \theta_k\big)$$

に置き換えて

$$\sqrt{N}\big[c(\hat{\theta}) - c(\theta)\big] \xrightarrow{d} \mathrm{Normal}\big(0,\ C(\theta)\, V\, C(\theta)^\top\big). \tag{8.3}$$

補足：線形変換との関係

線形関数 $c(\theta) = R\theta$ ではヤコビ行列が R そのものになり、近似ではなく厳密に $\sqrt{N}(R\hat{\theta} - R\theta) \xrightarrow{d} \mathrm{Normal}(0, RVR^\top)$ となります。第7章の線形変換は、デルタ法の特別な場合です。この線形の場合も「デルタ法」と呼ぶことがありますが、一般には、デルタ法は主に非線形版を指します。

発展：$c'(\theta)=0$ のとき

$c'(\theta)=0$ のとき（上の例 (1) で $\mu=0$ のとき）、極限の分散は 0 となり、デルタ法は「$\sqrt{N}[c(\hat{\theta})-c(\theta)] \xrightarrow{p} 0$」としか教えてくれません。このときは $\sqrt{N}$ より速い倍率が必要です。例えば $\mu=0$ なら、$N\bar{w}^2=(\sqrt{N}\bar{w})^2 \xrightarrow{d} \sigma^2\chi^2_1$（分布収束版の連続写像定理）となります。ベクトル版でも、非退化な正規分布を得るには $C(\theta)$ のランクが Q であることが必要です。

要点

- デルタ法：$\sqrt{N}(\hat{\theta}-\theta) \xrightarrow{d} \text{Normal}(0,V)$ なら、$\sqrt{N}[c(\hat{\theta})-c(\theta)] \xrightarrow{d} \text{Normal}(0,c'(\theta)^2V)$。
- 直感は接線による近似です。$\hat{\theta}$ は θ の近くにしか現れないので、c は直線とみなせ、正規分布は傾き倍されます。
- 証明は、平均値の定理、連続写像定理（$c'(\bar{\theta}) \xrightarrow{p} c'(\theta)$）、Slutsky の定理（掛け算）の組み合わせです。
- ベクトル版では、$c'(\theta)$ をヤコビ行列 $C(\theta)$ に置き換え、分散は $CVC^\top$ になります。

次の章へ　これまでの証明では、「0 に収束する余分な項」を何度も無視しました。次の章では、そのような項を簡潔に扱う記法 o_p, O_p を整理します。

大学院とのつながり　弾力性・限界効果・比率など、経済学で関心のある量はたいてい係数の非線形関数です。その標準誤差はデルタ法で計算します。

第9章

o_p・O_p 記法と周辺の用語

ここまでの流れ　証明の途中で、「0 に確率収束する余分な項」を何度も無視してきました（Slutsky の定理の帰結、デルタ法の証明など）。
この章の問い　そのような項を、いちいち文章で説明せずに、簡潔かつ正確に書く方法はないのでしょうか。
この章で分かること　「確率変数の大きさのオーダー（桁）」を表す記法 $o_p(\cdot)$・$O_p(\cdot)$、その計算規則、そして漸近理論でよく使われる用語（w.p.a.1、漸近サイズ、検定の一致性など）です。

9.1 定義

要点：o_p と O_p

- $x_N = o_p(1)$：$x_N \xrightarrow{p} 0$。
- $x_N = O_p(1)$（**確率的に有界**）：任意の $\varepsilon > 0$ に対して、ある定数 $b_\varepsilon < \infty$ と N_ε があり、$N \geq N_\varepsilon$ ならば $\Pr(|x_N| > b_\varepsilon) < \varepsilon$。
- 正の数列 a_N について、$x_N = o_p(a_N)$ とは $x_N/a_N = o_p(1)$、$x_N = O_p(a_N)$ とは $x_N/a_N = O_p(1)$。

数式の読み方

- $O_p(1)$ の読み方：「どんなに小さな確率 ε を指定されても、N によらない1つの壁 b_ε を立てて、x_N がその壁を越える確率を ε 未満に抑えられる」ということです。N が大きくなっても x_N の分布が無限遠に逃げていかない、という意味です。
- $O_p(a_N)$ は「大きさがせいぜい a_N くらい」、$o_p(a_N)$ は「a_N と比べて無視できるほど小さい」です。ふつうの数列のランダウの記号 $O(a_N)$、$o(a_N)$ の確率版です。

9.2 例

計算例：基本の例

1. **確率収束するものは** $O_p(1)$：$x_N \xrightarrow{p} a$ なら、$b_\varepsilon = |a| + 1$ とすれば $\Pr(|x_N| > |a| + 1) \le \Pr(|x_N - a| > 1) \to 0$。例：$\frac{1}{N}\sum x_i^2 = O_p(1)$。

2. **分布収束するものは** $O_p(1)$：例として $\sqrt{N}(\bar{w}_N - \mu) = O_p(1)$。

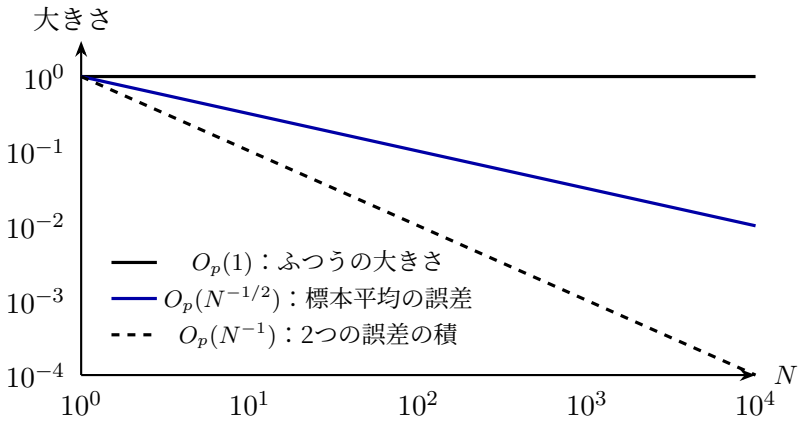
3. $\bar{w}_N - \mu = O_p(N^{-1/2})$：標本平均の誤差の大きさは $1/\sqrt{N}$ くらい、ということです。これはマルコフの不等式（$p = 2$）からも直接分かります。任意の $b > 0$ について

$$\begin{aligned} \Pr\bigl(\sqrt{N}|\bar{w}_N - \mu| > b\bigr) &\le \frac{N\,\mathrm{E}[(\bar{w}_N - \mu)^2]}{b^2} && \text{（マルコフの不等式）} \\ &= \frac{N \cdot \sigma^2/N}{b^2} = \frac{\sigma^2}{b^2} && \text{（}\mathrm{E}[(\bar{w}_N - \mu)^2] = \sigma^2/N\text{）} \end{aligned}$$

   なので、$b_\varepsilon = \sigma\sqrt{2/\varepsilon}$ とすればこの確率は $\varepsilon/2 < \varepsilon$ 以下になります。

4. $\sqrt{N}$**-一致性**：$\sqrt{N}(\hat{\theta} - \theta) = O_p(1)$、すなわち $\hat{\theta} - \theta = O_p(N^{-1/2})$。漸近正規な推定量はすべて $\sqrt{N}$-一致的です。

図 9.1　確率変数の大きさのオーダー：$O_p(1)$、$O_p(N^{-1/2})$、$O_p(N^{-1})$



大きさのオーダーを表す関数 1、$N^{-1/2}$、N^{-1} を、両対数のグラフに描いた（直線の傾きがオーダーの指数）。

図の読み方

1. 両対数のグラフでは、$N^{-\alpha}$ は傾き $-\alpha$ の直線になります。傾きが急なほど、速く 0 に潰れます。
2. 水平な線（$O_p(1)$）は潰れません。青の線（$O_p(N^{-1/2})$）は標本平均の誤差の大きさで、黒の破線（$O_p(N^{-1})$）はそのような誤差2つの積の大きさです。
3. 誤差の積は誤差1つよりずっと速く潰れます。これが、証明で「2つの誤差の積」が $o_p(1)$ として無視できる理由です。

9.3　計算規則

要点：計算規則

「o_p は小さい、O_p はふつうの大きさ」と考えれば自然です。

$$o_p(1)+o_p(1)=o_p(1), \qquad O_p(1)+o_p(1)=O_p(1),$$
$$O_p(1)\cdot o_p(1)=o_p(1), \qquad O_p(1)\cdot O_p(1)=O_p(1).$$

オーダー付きでも同様で、$O_p(a_N)\cdot O_p(b_N)=O_p(a_Nb_N)$、$O_p(a_N)\cdot o_p(b_N)=o_p(a_Nb_N)$ です。

証明（$O_p(1)\cdot o_p(1)=o_p(1)$）　$x_N=O_p(1)$、$y_N=o_p(1)$ とし、任意の $\delta>0$、$\varepsilon>0$ をとります。N が十分大きいとき $\Pr(|x_N|>b)<\varepsilon/2$ となる b を選びます。$|x_Ny_N|>\delta$ ならば、

「$|x_N| > b$」または「$|y_N| > \delta/b$」のどちらかが起きているので

$$\begin{aligned}\Pr(|x_N y_N| > \delta) &\le \Pr(|x_N| > b) + \Pr(|y_N| > \delta/b) && \text{（どちらかが起きている）}\\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon && \text{（N 十分大。$y_N = o_p(1)$ より第2項は $\varepsilon/2$ 未満）}\end{aligned}$$

です。ε は任意なので $\Pr(|x_N y_N| > \delta) \to 0$、すなわち $x_N y_N = o_p(1)$ です。

第5章の OLS の漸近正規性の証明の「$(\bar{x} - \mu_x) \cdot \frac{1}{\sqrt{N}} \sum u_i$」は、まさに $o_p(1) \cdot O_p(1) = o_p(1)$ でした。Slutsky の定理の帰結2は、この規則の言い換えです。

補足：つまずきポイント

これらは「等式」に見えますが、実際は「左辺の性質を持つものは、右辺の性質を持つ」という**一方向**の主張です。例えば $O_p(1) + o_p(1) = O_p(1)$ を右から左に読んで、「$O_p(1)$ のものは必ず $O_p(1) + o_p(1)$ に分解できる」と使うのは意味がありません。また $o_p(1)$ は $O_p(1)$ でもあります（$o_p(1) = O_p(1)$ は正しいが、$O_p(1) = o_p(1)$ は誤り）。

計算例：証明での使われ方の例

標本の共分散行列の推定量の一致性の証明では、次のような計算が登場します。平均 0 の IID 変数の標本平均は $O_p(1/\sqrt{N})$（上の例3）、$\sqrt{N}$-一致的な推定量の誤差も $O_p(1/\sqrt{N})$（例4）です。この2つの積は規則より $O_p(1/\sqrt{N}) \cdot O_p(1/\sqrt{N}) = O_p(1/N)$ で、とくに $o_p(1)$ です。また $O_p(1) \cdot O_p(1/N) = o_p(1)$ も、同じ規則から従います。
もう1つの使われ方は、$\sqrt{N}$-同値性です。2つの推定量 $\hat{\beta}$ と β^* の差が $\sqrt{N}(\hat{\beta} - \beta^*) = o_p(1)$ なら、$\sqrt{N}(\hat{\beta} - \beta) = \sqrt{N}(\beta^* - \beta) + o_p(1)$ なので、Slutsky の定理の帰結1より、両者は同じ漸近分布を持ちます。

9.4 その他の用語

w.p.a.1（with probability approaching one）　事象 E_N について $\Pr(E_N) \to 1$ となることです。例えば、$\hat{A} = \frac{1}{N} \sum x_i^\top x_i \xrightarrow{p} A$（$A$ は正則）なら、行列式は要素の連続関数なので $\det \hat{A} \xrightarrow{p} \det A \neq 0$ です。したがって

$$\begin{aligned}\Pr(\det \hat{A} \neq 0) &\ge \Pr\big(|\det \hat{A} - \det A| < |\det A|\big) && \text{（右の事象が起きれば } \det \hat{A} \neq 0\text{）}\\ &\to 1 && \text{（} \det \hat{A} \xrightarrow{p} \det A\text{）}\end{aligned}$$

つまり「$\hat{A}$ の逆行列は w.p.a.1 で存在する」と言えます。

漸近的効率性　同じパラメータの2つの漸近正規な推定量で、漸近分散の差 $D - V$ が半正定値なら、漸近分散 V の推定量の方が**効率的**です。

漸近サイズ　H_0 が正しいときに棄却する確率の極限 $\lim_{N\to\infty} \Pr(\text{棄却} \mid H_0)$ です。「$|t| > 1.96$ なら棄却」の検定の漸近サイズは 5% です（第6章）。

検定の一致性　H_1 が正しいとき、棄却する確率が 1 に近づくことです。例えば $\theta \neq b$ のとき、$t = \sqrt{N}(\hat{\theta} - b)/\sqrt{\hat{V}}$ の分子は

$$\sqrt{N}(\hat{\theta} - b) = \sqrt{N}(\theta - b) + O_p(1)$$

となります（$\sqrt{N}(\hat{\theta} - \theta) = O_p(1)$ を使った）。$\sqrt{N}(\theta - b)$ は $N \to \infty$ で発散するので、$|t|$ は確率的に無限大に発散し、$\Pr(|t| > 1.96) \to 1$ です。

要点

- $o_p(1)$ は「0 に確率収束」、$O_p(1)$ は「確率的に有界」。$O_p(a_N)$ は「せいぜい a_N くらいの大きさ」、$o_p(a_N)$ は「a_N に比べて無視できる」です。
- 確率収束するもの・分布収束するものは $O_p(1)$。標本平均の誤差は $O_p(N^{-1/2})$ で、漸近正規な推定量は $\sqrt{N}$-一致的です。
- 計算規則は「$O_p(1) \cdot o_p(1) = o_p(1)$」「$O_p(1) + o_p(1) = O_p(1)$」など。ただし等号は一方向の主張です。
- 周辺の用語：w.p.a.1（確率 1 に近づく）、漸近サイズ、検定の一致性、漸近的効率性。

次の章へ　道具はそろいました。最後の章で、道具の使い分けを一覧にまとめ、練習問題で理解を確かめます。

大学院とのつながり　大学院の教科書や論文では、$o_p(1)$ や $O_p(N^{-1/2})$ を使った式変形が当たり前に使われます。この記法に慣れておくと、証明の流れがずっと追いやすくなります。

第10章

総まとめと練習問題

ここまでの流れ　確率収束、大数の法則、連続写像定理、分布収束、中心極限定理、Slutsky の定理、二次形式、デルタ法、o_p・O_p 記法と、道具を1つずつ作ってきました。

この章の問い　これらの道具を、どの場面でどう使い分ければよいのでしょうか。理解は身についたでしょうか。

この章で分かること　道具の一覧、OLS の推論の各ステップと道具の対応、そして練習問題とその解答です。

10.1　道具箱の一覧

道具	内容	章
チェビシェフの不等式	$\Pr(\lvert Z - \mathrm{E}Z\rvert \geq \varepsilon) \leq \mathrm{Var}(Z)/\varepsilon^2$。確率収束を示す基本道具	2
大数の法則	IID 標本の標本平均 $\xrightarrow{p}$ 期待値	2
連続写像定理	g が連続なら、$x_N \xrightarrow{p} c \Rightarrow g(x_N) \xrightarrow{p} g(c)$。分布収束版は $x_N \xrightarrow{d} x \Rightarrow g(x_N) \xrightarrow{d} g(x)$	2, 4
中心極限定理	$\sqrt{N}(\bar{w}_N - \mu) \xrightarrow{d} \mathrm{Normal}(0, \sigma^2)$（ベクトル版は分散共分散行列 B）	4
Slutsky の定理	$x_N \xrightarrow{d} x$、$y_N \xrightarrow{p} c$（定数）なら和・積・商の極限は x と c で計算	5
線形変換・二次形式	$Rz_N \xrightarrow{d} \mathrm{Normal}(0, RVR^\top)$、$z_N^\top V^{-1} z_N \xrightarrow{d} \chi^2_K$	7
デルタ法	$\sqrt{N}[c(\hat{\theta}) - c(\theta)] \xrightarrow{d} \mathrm{Normal}(0, CVC^\top)$	8
o_p・O_p 記法	無視できる項と、ふつうの大きさの項を区別する	9

10.2 OLS の推論の各ステップと使う道具

式	使う道具	章
$\frac{1}{N}\sum x_i^\top x_i \xrightarrow{p} A$	大数の法則	2, 3
$\left[\frac{1}{N}\sum x_i^\top x_i\right]^{-1} \xrightarrow{p} A^{-1}$	連続写像定理、w.p.a.1	2, 9
$\frac{1}{\sqrt{N}}\sum x_i^\top u_i \xrightarrow{d} \text{Normal}(0, B)$	ベクトル版 CLT（$\text{E}(x_i^\top u_i) = 0$）	4
$\sqrt{N}(\hat{\beta} - \beta) \xrightarrow{d} \text{Normal}(0, A^{-1}BA^{-1})$	Slutsky の定理、線形変換	5, 7
$\hat{V} = \hat{A}^{-1}\hat{B}\hat{A}^{-1}$ が V の一致推定量	大数の法則、連続写像定理	6
$t \xrightarrow{d} \text{Normal}(0, 1)$	漸近正規性と $\hat{V} \xrightarrow{p} V$ の組み合わせ	6
$W \xrightarrow{d} \chi^2_Q$	二次形式	7

要点：証明を読むときのチェックリスト

1. 推定量が「真の値 + 標本平均の形の項」に分解されているか。
2. 各標本平均に、大数の法則（確率収束）と中心極限定理（分布収束）のどちらを使っているか。
3. 確率収束する部分が極限で定数（かつ、割り算なら 0 でない、逆行列なら正則）になっているか。
4. 部品の組み立てに、連続写像定理と Slutsky の定理のどちらを使っているか。

10.3 練習問題

問1 w_i が IID で $\text{Var}(w_i) = 4$ のとき、チェビシェフの不等式で $\Pr(|\bar{w}_N - \mu| \geq 0.1)$ の上限を、$N = 10{,}000$ について求めよ。

問2 $\hat{\theta}_N \xrightarrow{p} \theta > 0$ のとき、$\log \hat{\theta}_N$ と $1/\hat{\theta}_N$ はそれぞれ何に確率収束するか。どの定理のどの条件を使ったかも述べよ。

問3 Y_i が IID で $\text{E}(Y_i) = 31$、$\text{Var}(Y_i) = 59$ のとき、$N = 400$ の標本平均について $\Pr(|\bar{Y} - 31| \leq c) \approx 0.95$ となる c を、CLT を使って求めよ。

問4 単回帰で $\text{Cov}(x, u) = 0.5$、$\text{Var}(x) = 2$ のとき、$\text{plim}\, \hat{\beta}_2$ を β_2 を使って表せ。

問5 $\sqrt{N}(\hat{\theta} - \theta) \xrightarrow{d} \text{Normal}(0, V)$、$\theta \neq 0$ のとき、$\sqrt{N}(1/\hat{\theta} - 1/\theta)$ の漸近分布をデルタ法

で求めよ。

問6 $x_N = O_p(1)$、$y_N = o_p(1)$ のとき、$x_N^2 y_N + y_N$ は $o_p(1)$ か $O_p(1)$ か。

問7 $\hat{\theta}_N = w_1$（1人目のデータ）が μ の一致推定量でないことを、$w_i \sim \text{Normal}(\mu, 1)$ の場合に $\Pr(|\hat{\theta}_N - \mu| > 1)$ を計算して示せ。

問8 $Q = 1$（制約1本）のとき、Wald 統計量 W が t 統計量の2乗に等しいことを、$R = e_k^\top$（k 番目だけ 1 の行ベクトル）、$r = b$ として示せ。

10.4 解答

問1

$$\begin{aligned} \text{Var}(\bar{w}_N) &= \frac{4}{10000} = 0.0004 && (\text{Var}(\bar{w}_N) = \sigma^2/N) \\ \Pr(|\bar{w}_N - \mu| \geq 0.1) &\leq \frac{0.0004}{0.1^2} = 0.04 && (\text{チェビシェフの不等式}) \end{aligned}$$

上限は 0.04 です。

問2　$\log \hat{\theta}_N \xrightarrow{p} \log\theta$、$1/\hat{\theta}_N \xrightarrow{p} 1/\theta$。**連続写像定理**を使います。$\log x$ と $1/x$ は、どちらも $x = \theta > 0$ で連続なので、定理の条件「g が極限点で連続」を満たします。

問3　CLT より $\bar{Y}$ は近似的に $\text{Normal}(31,\, 59/400)$ に従います。標準偏差は $\sqrt{59/400} \approx 0.384$ です。標準化した値 $(\bar{Y} - 31)/0.384$ は近似的に $\text{Normal}(0,1)$ で、$\Pr(|Z| \leq 1.96) = 0.95$ なので

$$c = 1.96 \times 0.384 \approx 0.75.$$

問4　式 (3.2) より $\text{plim}\, \hat{\beta}_2 = \beta_2 + \dfrac{0.5}{2} = \beta_2 + 0.25$。$\text{Cov}(x, u) \neq 0$ なので、一致推定量ではありません。

問5　$c(\theta) = 1/\theta$、$c'(\theta) = -1/\theta^2$（$\theta \neq 0$ の近くで連続）なので、デルタ法より

$$\sqrt{N}\Big(\frac{1}{\hat{\theta}} - \frac{1}{\theta}\Big) \xrightarrow{d} \text{Normal}\Big(0,\, \Big(\frac{-1}{\theta^2}\Big)^2 V\Big) = \text{Normal}\Big(0,\, \frac{V}{\theta^4}\Big).$$

問6

$$x_N^2 = O_p(1) \cdot O_p(1) = O_p(1)$$
$$x_N^2 y_N = O_p(1) \cdot o_p(1) = o_p(1)$$
$$x_N^2 y_N + y_N = o_p(1) + o_p(1) = o_p(1)$$

したがって $o_p(1)$ です（$o_p(1)$ は $O_p(1)$ でもあるので、$O_p(1)$ とも言えますが、$o_p(1)$ の方がより強い主張です）。

問7　$w_1 - \mu \sim \text{Normal}(0,1)$ なので、N によらず

$$\Pr(|w_1 - \mu| > 1) = 2\big(1 - \Phi(1)\big) \approx 0.317$$

です。$N \to \infty$ でも 0 に近づかないので、確率収束の定義を満たしません。したがって、$\hat{\theta}_N = w_1$ は μ の一致推定量ではありません（不偏ではあります）。

問8　$Q = 1$ では、$R\hat{\theta} - r = \hat{\theta}_k - b$（スカラー）、$R(\hat{V}/N)R^\top = \hat{V}_{kk}/N = \text{se}(\hat{\theta}_k)^2$（スカラー）です。よって

$$W = \frac{(\hat{\theta}_k - b)^2}{\text{se}(\hat{\theta}_k)^2} = t_k^2.$$

χ_1^2 の 0.95 分位点は $1.96^2 \approx 3.84$ なので、5% 水準の Wald 検定と両側 t 検定は同じ結論になります。

要点

- 漸近理論の道具は、大数の法則（確率収束）、中心極限定理（分布収束）、連続写像定理と Slutsky の定理（組み立て）、二次形式（複数制約）、デルタ法（非線形関数）、o_p/O_p 記法（簡潔な記述）です。
- OLS の一致性・漸近正規性・t 検定・Wald 検定は、すべて「分解 $\to$ 標本平均に法則を当てはめる $\to$ 部品を組み立てる」の型で証明できます。

次へ　道具がそろいました。次は、条件付き期待値と線形射影、識別、OLS・GLS、IV・2SLS・GMM といった大学院レベルの話題を、この道具で読み進めます。

大学院とのつながり　推定量の性質は、ほぼすべて「一致性」「漸近正規性」「漸近分散の一致推定」の3点で述べられます。3段階の型に当てはめて読めば、証明の見通しが立ちます。