第３章：漸近統計学

Jonathan Roth

数理計量経済学 I
ブラウン大学

アウトライン

1. 概要
2. 大数の法則、中心極限定理、連続写像定理
3. 漸近理論の実践への応用

モチベーション

- これまでに、標本平均 $\hat{\mu}$ が分散既知の正規分布に従う場合に、母平均に関する仮説検定を行う方法を見てきました。
- この状況は、$Y_i \sim \mathrm{N}(\mu, \sigma^2)$ かつ σ が既知である場合に生じます。
- しかし、このような状況は稀です… より一般的な場合、どのように「推論」を行えばよいでしょうか？

モチベーション

- これまでに、標本平均 $\hat{\mu}$ が分散既知の正規分布に従う場合に、母平均に関する仮説検定を行う方法を見てきました。
- この状況は、$Y_i \sim \mathrm{N}(\mu, \sigma^2)$ かつ σ が既知である場合に生じます。
- しかし、このような状況は稀です… より一般的な場合、どのように「推論」を行えばよいでしょうか？
- 幸いなことに、サンプルサイズが大きい場合、標本平均が正規分布に従うという仮定は非常に良い近似になります。

モチベーション

- これまでに、標本平均 $\hat{\mu}$ が分散既知の正規分布に従う場合に、母平均に関する仮説検定を行う方法を見てきました。
- この状況は、$Y_i \sim \mathrm{N}(\mu, \sigma^2)$ かつ σ が既知である場合に生じます。
- しかし、このような状況は稀です... より一般的な場合、どのように「推論」を行えばよいでしょうか？
- 幸いなことに、サンプルサイズが大きい場合、標本平均が正規分布に従うという仮定は非常に良い近似になります。
- 何をもって「良い近似」とするかは、漸近統計学（asymptotic statistics）によって定式化されます。これは $N \to \infty$ としたときの極限における $\hat{\mu}$ の分布を考えます。

重要な結果の概要

- 大数の法則（Law of Large Numbers, LLN）は、N が大きいとき、$\hat{\mu}$ は非常に高い確率で μ に近くなることを示します。

重要な結果の概要

- 大数の法則（Law of Large Numbers, LLN）は、N が大きいとき、$\hat{\mu}$ は非常に高い確率で μ に近くなることを示します。
- 中心極限定理（Central Limit Theorem, CLT）は、N が大きいとき、$\hat{\mu}$ の分布は近似的に平均 μ、分散 σ^2/N の正規分布に従うことを示します。

重要な結果の概要

- 大数の法則（Law of Large Numbers, LLN）は、N が大きいとき、$\hat{\mu}$ は非常に高い確率で μ に近くなることを示します。
- 中心極限定理（Central Limit Theorem, CLT）は、N が大きいとき、$\hat{\mu}$ の分布は近似的に平均 μ、分散 σ^2/N の正規分布に従うことを示します。
- 連続写像定理（Continuous Mapping Theorem, CMT）は、N が大きいとき、$\hat{\mu}$ の連続関数、例えば $g(\hat{\mu})$ も $g(\mu)$ に近くなることを示します。

アウトライン

1. 概要 ✓
2. LLN, CLT, CMT
3. 漸近理論の実践への応用

確率収束

- 直感的には、確率変数 X_N が x に確率収束（**converge in probability**）するとは、N が大きいときに X_N が x に「近い」確率がほぼ 1 になることを意味します。

確率収束

- 直感的には、確率変数 X_N が x に確率収束（**converge in probability**）するとは、N が大きいときに X_N が x に「近い」確率がほぼ 1 になることを意味します。

- 形式的には、すべての $\varepsilon > 0$ に対して以下が成り立つとき、X_N は x に確率収束するといい、$X_N \to_p x$ または $plim X_N = x$ と表記します：

$$P(|X_N - x| > \varepsilon) \to 0$$

確率収束

- 直感的には、確率変数 X_N が x に確率収束（**converge in probability**）するとは、N が大きいときに X_N が x に「近い」確率がほぼ 1 になることを意味します。

- 形式的には、すべての $\varepsilon > 0$ に対して以下が成り立つとき、X_N は x に確率収束するといい、$X_N \to_p x$ または $plim X_N = x$ と表記します：

$$P(|X_N - x| > \varepsilon) \to 0$$

- 定数 x に対して $X_N \to_p x$ であるとき、X_N は x に対して一致性（**consistent**）があるといいます。

確率収束

- 直感的には、確率変数 X_N が x に確率収束（**converge in probability**）するとは、N が大きいときに X_N が x に「近い」確率がほぼ 1 になることを意味します。

- 形式的には、すべての $\varepsilon > 0$ に対して以下が成り立つとき、X_N は x に確率収束するといい、$X_N \to_p x$ または $plim X_N = x$ と表記します：

$$P(|X_N - x| > \varepsilon) \to 0$$

- 定数 x に対して $X_N \to_p x$ であるとき、X_N は x に対して一致性（**consistent**）があるといいます。

- 通常 x は定数ですが、同様の定義で確率変数 X への収束（$X_N \to X$）を言うこともあります。

確率収束（続き）

- 有用な事実：もし $E[(X_N - x)^2] \to 0$ ならば、$X_N \to_p x$ である。

確率収束（続き）

- 有用な事実：もし $E[(X_N - x)^2] \to 0$ ならば、$X_N \to_p x$ である。
- 証明（理解は必須ではありません）：
  反復期待値の法則より：

$$
\begin{aligned}
E[(X_N - x)^2] = & P(|X_N - x| > \varepsilon) E[(X_N - x)^2 \,|\, |X_N - x| > \varepsilon] + \\
& P(|X_N - x| \leq \varepsilon) E[(X_N - x)^2 \,|\, |X_N - x| \leq \varepsilon]
\end{aligned}
$$

確率収束（続き）

- 有用な事実：もし $E[(X_N - x)^2] \to 0$ ならば、$X_N \to_p x$ である。
- 証明（理解は必須ではありません）：
  反復期待値の法則より：

$$
\begin{aligned}
E[(X_N - x)^2] =& P(|X_N - x| > \varepsilon) E[(X_N - x)^2 | |X_N - x| > \varepsilon] + \\
& P(|X_N - x| \leq \varepsilon) E[(X_N - x)^2 | |X_N - x| \leq \varepsilon] \\
\geq & P(|X_N - x| > \varepsilon) \varepsilon^2 + 0
\end{aligned}
$$

確率収束（続き）

- 有用な事実：もし $E[(X_N - x)^2] \to 0$ ならば、$X_N \to_p x$ である。
- 証明（理解は必須ではありません）：
  反復期待値の法則より：

$$\begin{aligned} E[(X_N - x)^2] =& P(|X_N - x| > \varepsilon) E[(X_N - x)^2 || X_N - x| > \varepsilon] + \\ & P(|X_N - x| \leq \varepsilon) E[(X_N - x)^2 || X_N - x| \leq \varepsilon] \\ \geq & P(|X_N - x| > \varepsilon) \varepsilon^2 + 0 \end{aligned}$$

  これは次を意味します：

$$P(|X_N - x| > \varepsilon) \leq E[(X_N - x)^2]/\varepsilon^2 \text{（チェビシェフの不等式）}$$

確率収束（続き）

- 有用な事実：もし $E[(X_N - x)^2] \to 0$ ならば、$X_N \to_p x$ である。
- 証明（理解は必須ではありません）：
反復期待値の法則より：

$$\begin{aligned} E[(X_N - x)^2] =& P(|X_N - x| > \varepsilon) E[(X_N - x)^2 || X_N - x| > \varepsilon] + \\ & P(|X_N - x| \leq \varepsilon) E[(X_N - x)^2 || X_N - x| \leq \varepsilon] \\ \geq & P(|X_N - x| > \varepsilon)\varepsilon^2 + 0 \end{aligned}$$

これは次を意味します：

$$P(|X_N - x| > \varepsilon) \leq E[(X_N - x)^2]/\varepsilon^2 \text{（チェビシェフの不等式）}$$

したがって、$E[(X_N - x)^2] \to 0$ ならば $P(|X_N - x| > \varepsilon) \to 0$ となります。

大数の法則

- 大数の法則。$Y_1, ..., Y_N$ が $Var(Y_i) = \sigma^2 < \infty$ の分布から独立同一（iid）に抽出されているとします。このとき：

$$\hat{\mu}_N = \frac{1}{N}\sum_{i=1}^{N} Y_i \rightarrow_p \mu = E[Y_i]$$

- 言葉で言えば：サンプルサイズが大きくなるにつれて、標本平均は高い確率で母平均に近くなります。

大数の法則

- 大数の法則。$Y_1, ..., Y_N$ が $Var(Y_i) = \sigma^2 < \infty$ の分布から独立同一（*iid*）に抽出されているとします。このとき：

$$\hat{\mu}_N = \frac{1}{N}\sum_{i=1}^{N} Y_i \rightarrow_p \mu = E[Y_i]$$

- 言葉で言えば：サンプルサイズが大きくなるにつれて、標本平均は高い確率で母平均に近くなります。

- 証明：前章で $E[\hat{\mu}_N] = \mu$ および $Var(\hat{\mu}_N) = \sigma^2/N$ であることを見ました。
  したがって：

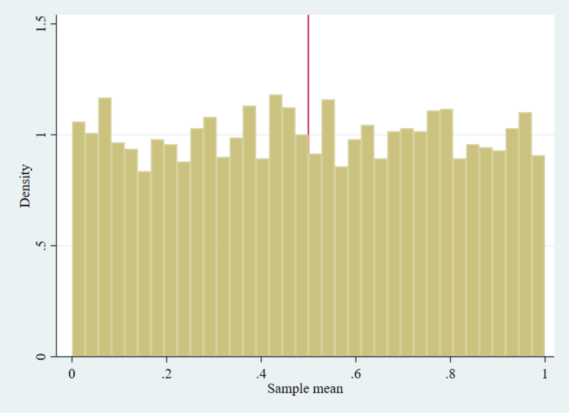
$$Var(\hat{\mu}_N) = E[(\hat{\mu}_N - \mu)^2] = \sigma^2/N \rightarrow 0$$

  ゆえに、上記の「有用な事実」により $\hat{\mu}_N \rightarrow_p \mu$ となります。

大数の法則のイラスト

$Z_i \sim \mathrm{U}(0,1)$ のときの $\frac{1}{N}\sum_i Z_i$ の分布と平均、$\mathbf{N} = \mathbf{1}$

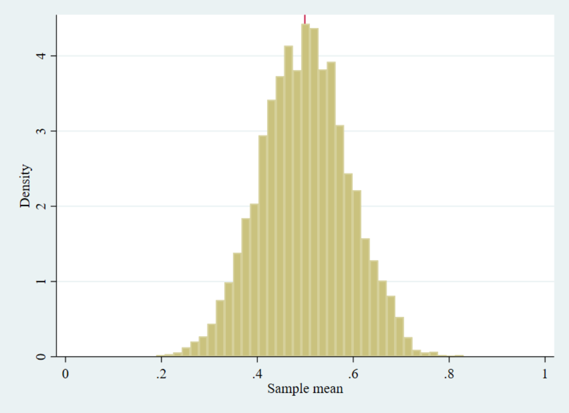
```
matrix sims=J(5000,1,.)
foreach N in 1 10 100 1000 {
    forval j=1/5000 {
        qui {
            clear
            set obs `N'
            gen X = runiform()
            summ X
            matrix sims[`j',1]=r(mean)
        }
    }
    clear
    svmat sims
    summ sims
    local mean = r(mean)
    hist sims, xlabel(0(0.2)1) xline(`mean')
    graph export sims`N'_2.png, replace
}
```



大数の法則のイラスト

$Z_i \sim \mathrm{U}(0,1)$ のときの $\frac{1}{N}\sum_i Z_i$ の分布と平均、$\mathbf{N = 10}$

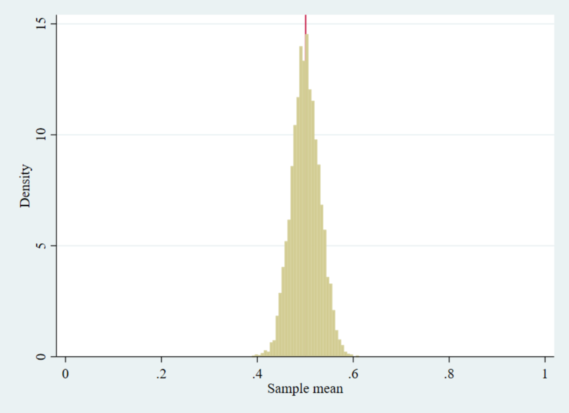
```
matrix sims=J(5000,1,.)
foreach N in 1 10 100 1000 {
    forval j=1/5000 {
        qui {
            clear
            set obs `N'
            gen X = runiform()
            summ X
            matrix sims[`j',1]=r(mean)
        }
    }
    clear
    svmat sims
    summ sims
    local mean = r(mean)
    hist sims, xlabel(0(0.2)1) xline(`mean')
    graph export sims`N'_2.png, replace
}
```



大数の法則のイラスト

$Z_i \sim \mathrm{U}(0,1)$ のときの $\frac{1}{N}\sum_i Z_i$ の分布と平均、$\mathbf{N = 100}$

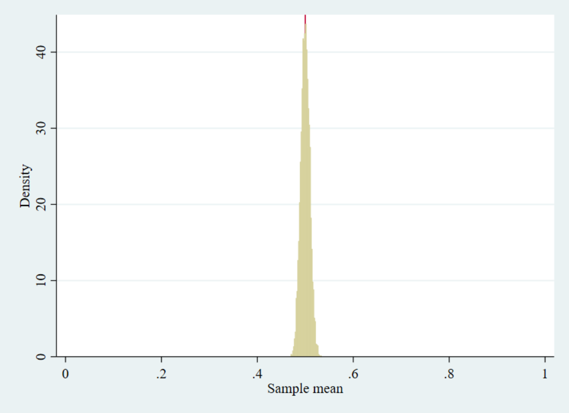
```
matrix sims=J(5000,1,.)
foreach N in 1 10 100 1000 {
    forval j=1/5000 {
        qui {
            clear
            set obs `N'
            gen X = runiform()
            summ X
            matrix sims[`j',1]=r(mean)
        }
    }
    clear
    svmat sims
    summ sims
    local mean = r(mean)
    hist sims, xlabel(0(0.2)1) xline(`mean')
    graph export sims`N'_2.png, replace
}
```



大数の法則のイラスト

$Z_i \sim \mathrm{U}(0,1)$ のときの $\frac{1}{N}\sum_i Z_i$ の分布と平均、**N** $=$ **1000**

```
matrix sims=J(5000,1,.)
foreach N in 1 10 100 1000 {
    forval j=1/5000 {
        qui {
            clear
            set obs `N'
            gen X = runiform()
            summ X
            matrix sims[`j',1]=r(mean)
        }
    }
    clear
    svmat sims
    summ sims
    local mean = r(mean)
    hist sims, xlabel(0(0.2)1) xline(`mean')
    graph export sims`N'_2.png, replace
}
```



分布収束

- シミュレーションにおいて、N が大きくなるにつれて $\hat{\mu}$ の分布が正規分布に近づいていることに気づいたかもしれません。
- 分布収束（**convergence in distribution**）という概念は、ある分布が別の分布に近づくことの意味を定式化したものです。

分布収束

- シミュレーションにおいて、N が大きくなるにつれて $\hat{\mu}$ の分布が正規分布に近づいていることに気づいたかもしれません。
- 分布収束（**convergence in distribution**）という概念は、ある分布が別の分布に近づくことの意味を定式化したものです。
- 定義：X_N の累積分布関数（CDF）が、連続型確率変数 X の累積分布関数に各点で収束するとき、すなわち：

$$F_{X_N}(x) \to F_X(x) \text{（すべての } x \text{ に対して）}$$

が成り立つとき、X_N は X に分布収束するといい、$X_N \to_d X$ または $X_N \Rightarrow X$ と表記します。

中心極限定理

- 中心極限定理（CLT）は、大標本において標本平均が近似的に正規分布に従うということを定式化したものです。

中心極限定理

- 中心極限定理（**CLT**）は、大標本において標本平均が近似的に正規分布に従うということを定式化したものです。
- 定理：$Y_1, ..., Y_N$ が平均 $\mu = E[Y_i]$、分散 $Var(Y_i) = \sigma^2 < \infty$ の分布から独立同一（iid）に抽出されているとします。このとき、標本平均 $\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} Y_i$ は以下を満たします：

$$\sqrt{N}(\hat{\mu} - \mu) \rightarrow_d N(0, \sigma^2)$$

中心極限定理

- 中心極限定理（CLT）は、大標本において標本平均が近似的に正規分布に従うということを定式化したものです。
- 定理：$Y_1, ..., Y_N$ が平均 $\mu = E[Y_i]$、分散 $Var(Y_i) = \sigma^2 < \infty$ の分布から独立同一（iid）に抽出されているとします。このとき、標本平均 $\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} Y_i$ は以下を満たします：

$$\sqrt{N}(\hat{\mu} - \mu) \rightarrow_d N(0, \sigma^2)$$

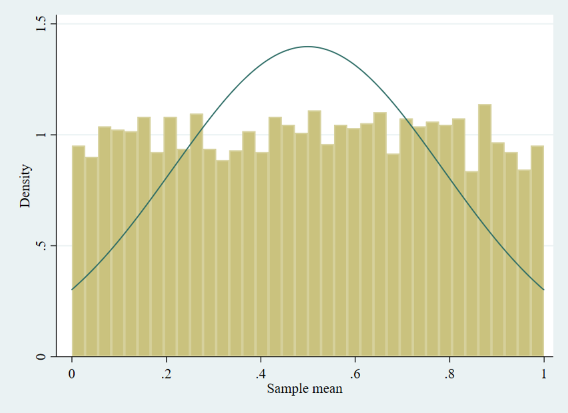
- 言葉で言えば、この定理は次のようなことを意味しています：
  1. 元の分布 Y_i がどのような分布（非正規分布でもよい）であっても、
  2. 十分に大きなサンプルの標本平均 $\hat{\mu} = \frac{1}{N}\sum_i Y_i$ の分布は、（近似的に）正規分布になります！

中心極限定理のイラスト

$\hat{\mu} = \frac{1}{N}\sum_i X_i$ の分布 vs. $\mathrm{N}(E[\hat{\mu}], Var(\hat{\mu}))$：$X_i \sim \mathrm{U}(0,1)$、$\mathbf{N} = \mathbf{1}$

```
clear all
set seed 42
set more off
set matsize 5000

matrix sims=J(5000,1,.)
foreach N in 1 2 5 10 {
    forval j=1/5000 {
        qui {
            clear
            set obs `N'
            gen X = runiform()
            summ X
            matrix sims[`j',1]=r(mean)
        }
    }
    clear
    svmat sims
    hist sims, normal xtitle("Sample mean")
    graph export sims`N'.png, replace
}
```

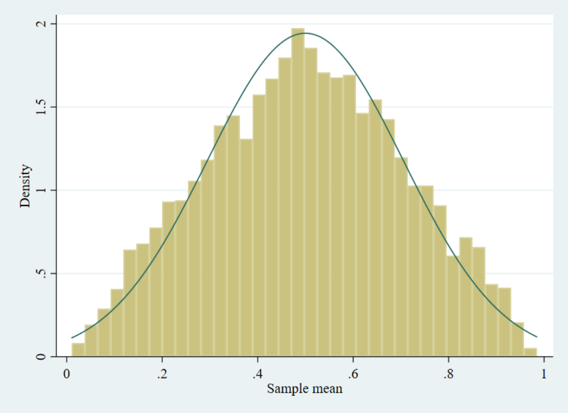


中心極限定理のイラスト

$\hat{\mu} = \frac{1}{N}\sum_i X_i$ の分布 vs. $\mathrm{N}(E[\hat{\mu}], Var(\hat{\mu}))$：$X_i \sim \mathrm{U}(0,1)$、**N** $=$ **2**

```
clear all
set seed 42
set more off
set matsize 5000

matrix sims=J(5000,1,.)
foreach N in 1 2 5 10 {
    forval j=1/5000 {
        qui {
            clear
            set obs `N'
            gen X = runiform()
            summ X
            matrix sims[`j',1]=r(mean)
        }
    }
    clear
    svmat sims
    hist sims, normal xtitle("Sample mean")
    graph export sims`N'.png, replace
}
```

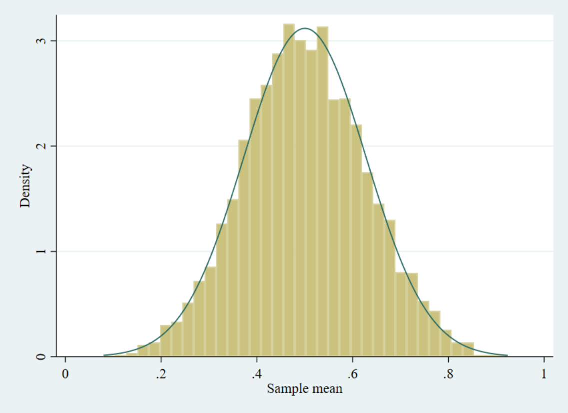


中心極限定理のイラスト

$\hat{\mu} = \frac{1}{N}\sum_i X_i$ の分布 vs. $\mathrm{N}(E[\hat{\mu}], Var(\hat{\mu}))$：$X_i \sim \mathrm{U}(0,1)$、$\mathbf{N} = \mathbf{5}$

```
clear all
set seed 42
set more off
set matsize 5000

matrix sims=J(5000,1,.)
foreach N in 1 2 5 10 {
    forval j=1/5000 {
        qui {
            clear
            set obs `N'
            gen X = runiform()
            summ X
            matrix sims[`j',1]=r(mean)
            }
        }
    clear
    svmat sims
    hist sims, normal xtitle("Sample mean")
    graph export sims`N'.png, replace
}
```

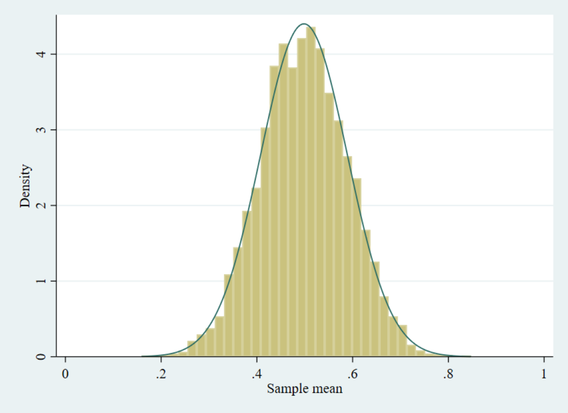


中心極限定理のイラスト

$\hat{\mu} = \frac{1}{N}\sum_i X_i$ の分布 vs. $\mathrm{N}(E[\hat{\mu}], Var(\hat{\mu}))$：$X_i \sim \mathrm{U}(0,1)$、$\mathbf{N} = \mathbf{10}$

```
clear all
set seed 42
set more off
set matsize 5000

matrix sims=J(5000,1,.)
foreach N in 1 2 5 10 {
    forval j=1/5000 {
        qui {
            clear
            set obs `N'
            gen X = runiform()
            summ X
            matrix sims[`j',1]=r(mean)
        }
    }
    clear
    svmat sims
    hist sims, normal xtitle("Sample mean")
    graph export sims`N'.png, replace
}
```



中心極限定理のイラスト II

https://www.youtube.com/watch?v=EvHiee7gs9Y

多変量版

- これまでの結果は、多変量の場合にも自然に拡張されます。

多変量版

- これまでの結果は、多変量の場合にも自然に拡張されます。
- ベクトル $\mathbf{X_N} \in \mathbb{R}^k$ について、各成分が確率収束するとき $\mathbf{X_N} \to_p \mathbf{x}$ といいます。

多変量版

- これまでの結果は、多変量の場合にも自然に拡張されます。
- ベクトル $\mathbf{X_N} \in \mathbb{R}^k$ について、各成分が確率収束するとき $\mathbf{X_N} \rightarrow_p \mathbf{x}$ といいます。
- **LLN**：平均 $\boldsymbol{\mu}$、有限な分散を持つ *iid* ベクトル $\mathbf{Y_1}, \ldots \mathbf{Y_N}$ の標本平均 $\hat{\boldsymbol{\mu}}_N$ について、$\hat{\boldsymbol{\mu}}_N \rightarrow_p \boldsymbol{\mu}$。

多変量版

- これまでの結果は、多変量の場合にも自然に拡張されます。
- ベクトル $\mathbf{X_N} \in \mathbb{R}^k$ について、各成分が確率収束するとき $\mathbf{X_N} \to_p \mathbf{x}$ といいます。
- **LLN**：平均 $\boldsymbol{\mu}$、有限な分散を持つ *iid* ベクトル $\mathbf{Y_1}, \ldots \mathbf{Y_N}$ の標本平均 $\hat{\boldsymbol{\mu}}_N$ について、$\hat{\boldsymbol{\mu}}_N \to_p \boldsymbol{\mu}$。
- ベクトル $\mathbf{X_N} \in \mathbb{R}^k$ について、$F_{\mathbf{X_N}}(\mathbf{x}) \to F_{\mathbf{X}}(\mathbf{x})$ が成り立つとき、$\mathbf{X_N} \to_d \mathbf{X}$ といいます。
- **CLT**：平均 $\boldsymbol{\mu}$、有限な分散 $\boldsymbol{\Sigma}$ を持つ *iid* ベクトル $\mathbf{Y_1}, \ldots \mathbf{Y_N}$ の標本平均 $\hat{\boldsymbol{\mu}}_N$ について、$\sqrt{N}(\hat{\boldsymbol{\mu}}_N - \boldsymbol{\mu}) \to_d \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma})$。

連続写像定理

- 時として、標本平均の関数（例：t 統計量は $\hat{\mu}$ と σ の関数です）に興味を持つことがあります。

連続写像定理

- 時として、標本平均の関数（例：t 統計量は $\hat{\mu}$ と σ の関数です）に興味を持つことがあります。
- 連続写像定理（Continuous Mapping Theorem, CMT）は、確率収束や分布収束する確率変数の連続関数について教えてくれます。

連続写像定理

- 時として、標本平均の関数（例：t 統計量は $\hat{\mu}$ と σ の関数です）に興味を持つことがあります。
- 連続写像定理（**Continuous Mapping Theorem, CMT**）は、確率収束や分布収束する確率変数の連続関数について教えてくれます。
- 定理：$g(\cdot)$ を連続関数とします。

  もし $X_N \to_p X$ ならば、$g(X_N) \to_p g(X)$。

連続写像定理

- 時として、標本平均の関数（例：t 統計量は $\hat{\mu}$ と σ の関数です）に興味を持つことがあります。
- 連続写像定理（**Continuous Mapping Theorem, CMT**）は、確率収束や分布収束する確率変数の連続関数について教えてくれます。
- 定理：$g(\cdot)$ を連続関数とします。

  もし $X_N \rightarrow_p X$ ならば、$g(X_N) \rightarrow_p g(X)$。

  もし $X_N \rightarrow_d X$ ならば、$g(X_N) \rightarrow_d g(X)$。

連続写像定理

- 時として、標本平均の関数（例：t 統計量は $\hat{\mu}$ と σ の関数です）に興味を持つことがあります。
- 連続写像定理（Continuous Mapping Theorem, CMT）は、確率収束や分布収束する確率変数の連続関数について教えてくれます。
- 定理：$g(\cdot)$ を連続関数とします。

  もし $X_N \rightarrow_p X$ ならば、$g(X_N) \rightarrow_p g(X)$。

  もし $X_N \rightarrow_d X$ ならば、$g(X_N) \rightarrow_d g(X)$。

  多変量版：$\mathbf{X_N} \rightarrow_p \mathbf{X}$ ならば $g(\mathbf{X_N}) \rightarrow_p g(\mathbf{X})$、$\mathbf{X_N} \rightarrow_d \mathbf{X}$ ならば $g(\mathbf{X_N}) \rightarrow_d g(\mathbf{X})$。

不偏分散の収束

- CMT の有用な応用例として、不偏分散の確率収束を示すことができます。

不偏分散の収束

- CMT の有用な応用例として、不偏分散の確率収束を示すことができます。
- $\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \hat{\mu})^2$ を Y_i の標本分散とします。

不偏分散の収束

- CMT の有用な応用例として、不偏分散の確率収束を示すことができます。
- $\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \hat{\mu})^2$ を Y_i の標本分散とします。
- 主張：$Y_1, ..., Y_N$ が iid であり $Var(Y_i^2)$ が有限であれば、$\hat{\sigma}^2 \to_p \sigma^2 = Var(Y_i)$。

不偏分散の収束

- CMT の有用な応用例として、不偏分散の確率収束を示すことができます。

- $\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \hat{\mu})^2$ を Y_i の標本分散とします。

- 主張：$Y_1, ..., Y_N$ が iid であり $Var(Y_i^2)$ が有限であれば、
$\hat{\sigma}^2 \rightarrow_p \sigma^2 = Var(Y_i)$。

- 証明：
標本分散は $\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N} Y_i^2 - \hat{\mu}^2$ と書けます。

不偏分散の収束

- CMT の有用な応用例として、不偏分散の確率収束を示すことができます。
- $\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \hat{\mu})^2$ を Y_i の標本分散とします。
- 主張：$Y_1, ..., Y_N$ が iid であり $Var(Y_i^2)$ が有限であれば、
$\hat{\sigma}^2 \to_p \sigma^2 = Var(Y_i)$。
- 証明：
標本分散は $\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N} Y_i^2 - \hat{\mu}^2$ と書けます。
第 1 項：LLN より、$\frac{1}{N}\sum_{i=1}^{N} Y_i^2 \to_p E[Y_i^2]$。

不偏分散の収束

- CMT の有用な応用例として、不偏分散の確率収束を示すことができます。

- $\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \hat{\mu})^2$ を Y_i の標本分散とします。

- 主張：$Y_1, ..., Y_N$ が iid であり $Var(Y_i^2)$ が有限であれば、$\hat{\sigma}^2 \to_p \sigma^2 = Var(Y_i)$。

- 証明：
  標本分散は $\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N} Y_i^2 - \hat{\mu}^2$ と書けます。
  第 1 項：LLN より、$\frac{1}{N}\sum_{i=1}^{N} Y_i^2 \to_p E[Y_i^2]$。
  第 2 項：LLN より、$\hat{\mu} \to_p \mu = E[Y_i]$。よって CMT より $\hat{\mu}^2 \to_p E[Y_i]^2$。

不偏分散の収束

- CMT の有用な応用例として、不偏分散の確率収束を示すことができます。

- $\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \hat{\mu})^2$ を Y_i の標本分散とします。

- 主張：$Y_1, ..., Y_N$ が iid であり $Var(Y_i^2)$ が有限であれば、$\hat{\sigma}^2 \rightarrow_p \sigma^2 = Var(Y_i)$。

- 証明：
  標本分散は $\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N} Y_i^2 - \hat{\mu}^2$ と書けます。
  第 1 項：LLN より、$\frac{1}{N}\sum_{i=1}^{N} Y_i^2 \rightarrow_p E[Y_i^2]$。
  第 2 項：LLN より、$\hat{\mu} \rightarrow_p \mu = E[Y_i]$。よって CMT より $\hat{\mu}^2 \rightarrow_p E[Y_i]^2$。
  したがって、再び CMT を用いると、
  $\frac{1}{N}\sum_{i=1}^{N} Y_i^2 - \hat{\mu}^2 \rightarrow_p E[Y_i^2] - E[Y_i]^2 = \sigma^2$。

スルツキーの定理

- スルツキーの定理（**Slutsky's lemma**）は、CMT の非常に有用な特殊ケースをまとめたものです。

スルツキーの定理

- スルツキーの定理（**Slutsky’s lemma**）は、CMT の非常に有用な特殊ケースをまとめたものです。
- 定数 c に対して $X_N \to_p c$ であり、$Y_N \to_d Y$ とします。このとき：
- $X_N + Y_N \to_d c + Y$

スルツキーの定理

- スルツキーの定理（**Slutsky's lemma**）は、CMT の非常に有用な特殊ケースをまとめたものです。
- 定数 c に対して $X_N \to_p c$ であり、$Y_N \to_d Y$ とします。このとき：
- $X_N + Y_N \to_d c + Y$
- $X_N Y_N \to_d cY$

スルツキーの定理

- スルツキーの定理（**Slutsky's lemma**）は、CMT の非常に有用な特殊ケースをまとめたものです。
- 定数 c に対して $X_N \to_p c$ であり、$Y_N \to_d Y$ とします。このとき：
- $X_N + Y_N \to_d c + Y$
- $X_N Y_N \to_d cY$
- もし $c \neq 0$ ならば、$Y_N / X_N \to_d Y / c$

スルツキーの定理

- スルツキーの定理（**Slutsky's lemma**）は、CMT の非常に有用な特殊ケースをまとめたものです。
- 定数 c に対して $X_N \rightarrow_p c$ であり、$Y_N \rightarrow_d Y$ とします。このとき：
- $X_N + Y_N \rightarrow_d c + Y$
- $X_N Y_N \rightarrow_d cY$
- もし $c \neq 0$ ならば、$Y_N / X_N \rightarrow_d Y/c$
- ベクトル値の確率変数についても同様の結果が成り立ちます。

漸近的な仮説検定

- $Y_i \sim \mathrm{N}(\mu, \sigma^2)$ のとき、帰無仮説 $H_0 : \mu = \mu_0$ の下で t 統計量 $\hat{t} = \frac{\hat{\mu} - \mu_0}{\sigma/\sqrt{N}} \sim \mathrm{N}(0, 1)$ となることを見ました。

漸近的な仮説検定

- $Y_i \sim \mathrm{N}(\mu, \sigma^2)$ のとき、帰無仮説 $H_0 : \mu = \mu_0$ の下で t 統計量 $\hat{t} = \frac{\hat{\mu} - \mu_0}{\sigma/\sqrt{N}} \sim \mathrm{N}(0, 1)$ となることを見ました。
- したがって、$Y_i \sim \mathrm{N}(\mu, \sigma^2)$ のときは、帰無仮説の下で $Pr(|\hat{t}| > 1.96) = 0.05$ でした。

漸近的な仮説検定

- $Y_i \sim \mathrm{N}(\mu, \sigma^2)$ のとき、帰無仮説 $H_0 : \mu = \mu_0$ の下で t 統計量 $\hat{t} = \frac{\hat{\mu}-\mu_0}{\sigma/\sqrt{N}} \sim \mathrm{N}(0,1)$ となることを見ました。
- したがって、$Y_i \sim \mathrm{N}(\mu, \sigma^2)$ のときは、帰無仮説の下で $Pr(|\hat{t}| > 1.96) = 0.05$ でした。
- ここで、Y_i が正規分布に従わず、分散も未知である場合を考えましょう。

漸近的な仮説検定

- $Y_i \sim \mathrm{N}(\mu, \sigma^2)$ のとき、帰無仮説 $H_0 : \mu = \mu_0$ の下で t 統計量 $\hat{t} = \frac{\hat{\mu} - \mu_0}{\sigma/\sqrt{N}} \sim \mathrm{N}(0, 1)$ となることを見ました。
- したがって、$Y_i \sim \mathrm{N}(\mu, \sigma^2)$ のときは、帰無仮説の下で $Pr(|\hat{t}| > 1.96) = 0.05$ でした。
- ここで、Y_i が正規分布に従わず、分散も未知である場合を考えましょう。
- CLT より、$\sqrt{N}(\hat{\mu} - \mu_0) \to_d N(0, \sigma^2)$。
  CMT および LLN より、$\hat{\sigma} \to_p \sigma$。

漸近的な仮説検定

- $Y_i \sim \mathrm{N}(\mu, \sigma^2)$ のとき、帰無仮説 $H_0 : \mu = \mu_0$ の下で t 統計量 $\hat{t} = \frac{\hat{\mu} - \mu_0}{\sigma/\sqrt{N}} \sim \mathrm{N}(0, 1)$ となることを見ました。
- したがって、$Y_i \sim \mathrm{N}(\mu, \sigma^2)$ のときは、帰無仮説の下で $Pr(|\hat{t}| > 1.96) = 0.05$ でした。
- ここで、Y_i が正規分布に従わず、分散も未知である場合を考えましょう。
- CLT より、$\sqrt{N}(\hat{\mu} - \mu_0) \to_d N(0, \sigma^2)$。
  CMT および LLN より、$\hat{\sigma} \to_p \sigma$。
- よってスルツキーの定理より、$\hat{t} = \dfrac{\hat{\mu} - \mu_0}{\hat{\sigma}/\sqrt{N}} \to_d N(0, 1)$。

漸近的な仮説検定

- $Y_i \sim \mathrm{N}(\mu, \sigma^2)$ のとき、帰無仮説 $H_0 : \mu = \mu_0$ の下で t 統計量 $\hat{t} = \frac{\hat{\mu}-\mu_0}{\sigma/\sqrt{N}} \sim \mathrm{N}(0,1)$ となることを見ました。
- したがって、$Y_i \sim \mathrm{N}(\mu, \sigma^2)$ のときは、帰無仮説の下で $Pr(|\hat{t}| > 1.96) = 0.05$ でした。
- ここで、Y_i が正規分布に従わず、分散も未知である場合を考えましょう。
- CLT より、$\sqrt{N}(\hat{\mu} - \mu_0) \to_d N(0, \sigma^2)$。
  CMT および LLN より、$\hat{\sigma} \to_p \sigma$。
- よってスルツキーの定理より、$\hat{t} = \dfrac{\hat{\mu} - \mu_0}{\hat{\sigma}/\sqrt{N}} \to_d N(0,1)$。
- ゆえに、漸近的に $Pr(|\hat{t}| > 1.96) \to 0.05$ となります。Y_i が非正規で $\hat{\sigma}$ が推定値であっても、以前と同じように仮説検定ができるのです！

漸近的な信頼区間

- 同様に、Y_i が正規分布で σ が既知のとき、信頼区間 $\hat{\mu} \pm 1.96\sigma/\sqrt{N}$ が真の μ を含む確率は 95% でした。

漸近的な信頼区間

- 同様に、Y_i が正規分布で σ が既知のとき、信頼区間 $\hat{\mu} \pm 1.96\sigma/\sqrt{N}$ が真の μ を含む確率は 95% でした。

- 同じように、Y_i が非正規で分散が未知の場合でも、$\hat{\mu} \pm 1.96\hat{\sigma}/\sqrt{N}$ が真の μ を含む確率は、N が大きくなるにつれて 95% に近づきます。

アウトライン

1. 概要 ✓
2. LLN, CLT, CMT ✓
3. 漸近理論の実践への応用

例：オレゴン健康保険実験

In 2008, a group of uninsured low-income adults in Oregon was selected by lottery to be given the chance to apply for Medicaid. This lottery provides an opportunity to gauge the effects of expanding access to public health insurance on the health care use, financial strain, and health of low-income adults using a randomized controlled design. In the year after random assignment, the treatment group selected by the lottery was about 25 percentage points more likely to have insurance than the control group that was not selected. We find that in this first year, the treatment group had substantively and statistically significantly higher health care utilization (including primary and preventive care as well as hospitalizations), lower out-of-pocket medical expenditures and medical debt (including fewer bills sent to collection), and better self-reported physical and mental health than the control group. *JEL* Codes: H51, H75, I1.

うつ病の結果に関する標本平均

	対照群（Control）	処置群（Treated）
平均	0.329	0.306
標準偏差	0.470	0.461
サンプルサイズ (N)	10426	13315

うつ病の結果に関する標本平均

	対照群（Control）	処置群（Treated）
平均	0.329	0.306
標準偏差	0.470	0.461
サンプルサイズ (N)	10426	13315

- 対照群における母平均の信頼区間（CI）を求めたいとします。

うつ病の結果に関する標本平均

	対照群（Control）	処置群（Treated）
平均	0.329	0.306
標準偏差	0.470	0.461
サンプルサイズ (N)	10426	13315

- 対照群における母平均の信頼区間（CI）を求めたいとします。
- 次のように計算できます：

$$\hat{\mu} \pm 1.96 \times \hat{\sigma}/\sqrt{N} =$$

うつ病の結果に関する標本平均

	対照群（Control）	処置群（Treated）
平均	0.329	0.306
標準偏差	0.470	0.461
サンプルサイズ (N)	10426	13315

- 対照群における母平均の信頼区間（CI）を求めたいとします。
- 次のように計算できます：

$$\hat{\mu} \pm 1.96 \times \hat{\sigma}/\sqrt{N} = 0.329 \pm 1.96 \times 0.470/\sqrt{10426} =$$

うつ病の結果に関する標本平均

	対照群（Control）	処置群（Treated）
平均	0.329	0.306
標準偏差	0.470	0.461
サンプルサイズ (N)	10426	13315

- 対照群における母平均の信頼区間（CI）を求めたいとします。
- 次のように計算できます：

$$\hat{\mu} \pm 1.96 \times \hat{\sigma}/\sqrt{N} = 0.329 \pm 1.96 \times 0.470/\sqrt{10426} = [0.319, 0.338]$$

うつ病の結果に関する標本平均

	対照群（Control）	処置群（Treated）
平均	0.329	0.306
標準偏差	0.470	0.461
サンプルサイズ (N)	10426	13315

- 対照群における母平均の信頼区間（CI）を求めたいとします。
- 次のように計算できます：

$$\hat{\mu} \pm 1.96 \times \hat{\sigma}/\sqrt{N} = 0.329 \pm 1.96 \times 0.470/\sqrt{10426} = [0.319, 0.338]$$

- 処置群についてはどうでしょうか？

うつ病の結果に関する標本平均

	対照群（Control）	処置群（Treated）
平均	0.329	0.306
標準偏差	0.470	0.461
サンプルサイズ (N)	10426	13315

- 対照群における母平均の信頼区間（CI）を求めたいとします。
- 次のように計算できます：

$$\hat{\mu} \pm 1.96 \times \hat{\sigma}/\sqrt{N} = 0.329 \pm 1.96 \times 0.470/\sqrt{10426} = [0.319, 0.338]$$

- 処置群についてはどうでしょうか？

$$\hat{\mu} \pm 1.96 \times \hat{\sigma}/\sqrt{N} =$$

うつ病の結果に関する標本平均

	対照群（Control）	処置群（Treated）
平均	0.329	0.306
標準偏差	0.470	0.461
サンプルサイズ (N)	10426	13315

- 対照群における母平均の信頼区間（CI）を求めたいとします。
- 次のように計算できます：

$$\hat{\mu} \pm 1.96 \times \hat{\sigma}/\sqrt{N} = 0.329 \pm 1.96 \times 0.470/\sqrt{10426} = [0.319, 0.338]$$

- 処置群についてはどうでしょうか？

$$\hat{\mu} \pm 1.96 \times \hat{\sigma}/\sqrt{N} = 0.306 \pm 1.96 \times 0.461/\sqrt{13315} =$$

うつ病の結果に関する標本平均

	対照群（Control）	処置群（Treated）
平均	0.329	0.306
標準偏差	0.470	0.461
サンプルサイズ (N)	10426	13315

- 対照群における母平均の信頼区間（CI）を求めたいとします。
- 次のように計算できます：

$$\hat{\mu} \pm 1.96 \times \hat{\sigma}/\sqrt{N} = 0.329 \pm 1.96 \times 0.470/\sqrt{10426} = [0.319, 0.338]$$

- 処置群についてはどうでしょうか？

$$\hat{\mu} \pm 1.96 \times \hat{\sigma}/\sqrt{N} = 0.306 \pm 1.96 \times 0.461/\sqrt{13315} = [0.298, 0.313]$$

実験における処置効果の信頼区間

- 以前示したように、実験における平均処置効果（ATE）は次のように与えられます：

$$\tau = E[Y_i(1) - Y_i(0)] = E[Y_i | D_i = 1] - E[Y_i | D_i = 0]$$

すなわち、処置群と対照群の母平均の差です。

実験における処置効果の信頼区間

- 以前示したように、実験における平均処置効果（ATE）は次のように与えられます：

$$\tau = E[Y_i(1) - Y_i(0)] = E[Y_i | D_i = 1] - E[Y_i | D_i = 0]$$

  すなわち、処置群と対照群の母平均の差です。

- この処置効果について、どのように信頼区間を形成（あるいは仮説検定を）すればよいでしょうか？

平均の差の期待値と分散

- $\bar{Y}_1 = \frac{1}{N_1}\sum_{i:D_i=1} Y_i$ を処置群の標本平均とします。
  $\bar{Y}_0 = \frac{1}{N_0}\sum_{i:D_i=0} Y_i$ を対照群の標本平均とします。

平均の差の期待値と分散

- $\bar{Y}_1 = \frac{1}{N_1} \sum_{i:D_i=1} Y_i$ を処置群の標本平均とします。
  $\bar{Y}_0 = \frac{1}{N_0} \sum_{i:D_i=0} Y_i$ を対照群の標本平均とします。
- $\bar{Y}_1, \bar{Y}_0$ はそれぞれ標本平均なので：

平均の差の期待値と分散

- $\bar{Y}_1 = \frac{1}{N_1}\sum_{i:D_i=1} Y_i$ を処置群の標本平均とします。
  $\bar{Y}_0 = \frac{1}{N_0}\sum_{i:D_i=0} Y_i$ を対照群の標本平均とします。

- $\bar{Y}_1, \bar{Y}_0$ はそれぞれ標本平均なので：

$$
\begin{aligned}
E[\bar{Y}_1] &= \mu_1, & Var(\bar{Y}_1) &= \sigma_1^2/N_1 \\
E[\bar{Y}_0] &= \mu_0, & Var(\bar{Y}_0) &= \sigma_0^2/N_0
\end{aligned}
$$

  ここで $\mu_d = E[Y_i \mid D_i = d]$、$\sigma_d^2 = Var(Y_i \mid D_i = d)$ です。

- $\hat{\tau} = \bar{Y}_1 - \bar{Y}_0$ とします。すると $E[\hat{\tau}] =$

平均の差の期待値と分散

- $\bar{Y}_1 = \frac{1}{N_1}\sum_{i:D_i=1} Y_i$ を処置群の標本平均とします。
  $\bar{Y}_0 = \frac{1}{N_0}\sum_{i:D_i=0} Y_i$ を対照群の標本平均とします。

- $\bar{Y}_1, \bar{Y}_0$ はそれぞれ標本平均なので：

$$E[\bar{Y}_1] = \mu_1, \qquad Var(\bar{Y}_1) = \sigma_1^2 / N_1$$
$$E[\bar{Y}_0] = \mu_0, \qquad Var(\bar{Y}_0) = \sigma_0^2 / N_0$$

  ここで $\mu_d = E[Y_i \mid D_i = d]$、$\sigma_d^2 = Var(Y_i \mid D_i = d)$ です。

- $\hat{\tau} = \bar{Y}_1 - \bar{Y}_0$ とします。すると $E[\hat{\tau}] = \mu_1 - \mu_0 = \tau$ であり：

$$Var(\hat{\tau}) =$$

平均の差の期待値と分散

- $\bar{Y}_1 = \frac{1}{N_1}\sum_{i:D_i=1} Y_i$ を処置群の標本平均とします。
  $\bar{Y}_0 = \frac{1}{N_0}\sum_{i:D_i=0} Y_i$ を対照群の標本平均とします。

- $\bar{Y}_1, \bar{Y}_0$ はそれぞれ標本平均なので：

$$E[\bar{Y}_1] = \mu_1, \qquad Var(\bar{Y}_1) = \sigma_1^2/N_1$$
$$E[\bar{Y}_0] = \mu_0, \qquad Var(\bar{Y}_0) = \sigma_0^2/N_0$$

  ここで $\mu_d = E[Y_i \mid D_i = d]$、$\sigma_d^2 = Var(Y_i \mid D_i = d)$ です。

- $\hat{\tau} = \bar{Y}_1 - \bar{Y}_0$ とします。すると $E[\hat{\tau}] = \mu_1 - \mu_0 = \tau$ であり：

$$\begin{aligned} Var(\hat{\tau}) &= \sigma_1^2/N_1 + \sigma_0^2/N_0 - 2Cov(\bar{Y}_1, \bar{Y}_0) \\ &= \end{aligned}$$

平均の差の期待値と分散

- $\bar{Y}_1 = \frac{1}{N_1}\sum_{i:D_i=1} Y_i$ を処置群の標本平均とします。
  $\bar{Y}_0 = \frac{1}{N_0}\sum_{i:D_i=0} Y_i$ を対照群の標本平均とします。

- $\bar{Y}_1, \bar{Y}_0$ はそれぞれ標本平均なので：

$$
\begin{aligned}
E[\bar{Y}_1] &= \mu_1, & \qquad Var(\bar{Y}_1) &= \sigma_1^2/N_1 \\
E[\bar{Y}_0] &= \mu_0, & \qquad Var(\bar{Y}_0) &= \sigma_0^2/N_0
\end{aligned}
$$

  ここで $\mu_d = E[Y_i \mid D_i = d]$、$\sigma_d^2 = Var(Y_i \mid D_i = d)$ です。

- $\hat{\tau} = \bar{Y}_1 - \bar{Y}_0$ とします。すると $E[\hat{\tau}] = \mu_1 - \mu_0 = \tau$ であり：

$$
\begin{aligned}
Var(\hat{\tau}) &= \sigma_1^2/N_1 + \sigma_0^2/N_0 - 2Cov(\bar{Y}_1, \bar{Y}_0) \\
&= \sigma_1^2/N_1 + \sigma_0^2/N_0
\end{aligned}
$$

  ここで、各サンプルが独立であるという事実から $Cov(\bar{Y}_1, \bar{Y}_0) = 0$ となります。

- 実験において次が成り立つことを示しました：

$$E[\hat{\tau}] = \tau \text{ および } Var(\hat{\tau}) = \sigma_1^2/N_1 + \sigma_0^2/N_0$$

  ここで $\hat{\tau}$ は処置群と対照群の標本平均の差です。

- もし $\hat{\tau}$ が正規分布に従う（かつ σ_1, σ_0 が既知である）と分かっていれば、次のような信頼区間を構築できます：

- 実験において次が成り立つことを示しました：

$$E[\hat{\tau}] = \tau \text{ および } Var(\hat{\tau}) = \sigma_1^2/N_1 + \sigma_0^2/N_0$$

  ここで $\hat{\tau}$ は処置群と対照群の標本平均の差です。

- もし $\hat{\tau}$ が正規分布に従う（かつ σ_1, σ_0 が既知である）と分かっていれば、次のような信頼区間を構築できます：

$$\hat{\tau} \pm 1.96\sqrt{\sigma_1^2/N_1 + \sigma_0^2/N_0}$$

- 実験において次が成り立つことを示しました：

$$E[\hat{\tau}] = \tau \text{ および } Var(\hat{\tau}) = \sigma_1^2/N_1 + \sigma_0^2/N_0$$

ここで $\hat{\tau}$ は処置群と対照群の標本平均の差です。

- もし $\hat{\tau}$ が正規分布に従う（かつ σ_1, σ_0 が既知である）と分かっていれば、次のような信頼区間を構築できます：

$$\hat{\tau} \pm 1.96\sqrt{\sigma_1^2/N_1 + \sigma_0^2/N_0}$$

- 標本平均の場合と同様に、$\hat{\tau}$ が正規分布に従うとは限りませんが、N が大きければ近似的に正規分布に従うことを示すことができます。これにより、推定された条件付き分散 $\hat{\sigma}_d^2$ を用いて以下の信頼区間を使用できます：

$$\hat{\tau} \pm 1.96\sqrt{\hat{\sigma}_1^2/N_1 + \hat{\sigma}_0^2/N_0}$$

漸近正規性の証明

- CLT より、$\sqrt{N_1}(\bar{Y}_1 - \mu_1) \rightarrow_d$

漸近正規性の証明

- CLT より、$\sqrt{N_1}(\bar{Y}_1 - \mu_1) \to_d N(0, \sigma_1^2)$。

漸近正規性の証明

- CLT より、$\sqrt{N_1}(\bar{Y}_1 - \mu_1) \to_d N(0, \sigma_1^2)$。
- LLN より、$\frac{N_1}{N} = \frac{1}{N}\sum_i D_i \to_p$

漸近正規性の証明

- CLT より、$\sqrt{N_1}(\bar{Y}_1 - \mu_1) \to_d N(0, \sigma_1^2)$。
- LLN より、$\frac{N_1}{N} = \frac{1}{N}\sum_i D_i \to_p E[D_i]$。

漸近正規性の証明

- CLT より、$\sqrt{N_1}(\bar{Y}_1 - \mu_1) \to_d N(0, \sigma_1^2)$。
- LLN より、$\frac{N_1}{N} = \frac{1}{N}\sum_i D_i \to_p E[D_i]$。
- したがって、連続写像定理を適用すると：

$$\sqrt{N}(\bar{Y}_1 - E[Y_i(1)]) =$$

漸近正規性の証明

- CLT より、$\sqrt{N_1}(\bar{Y}_1 - \mu_1) \to_d N(0, \sigma_1^2)$。
- LLN より、$\frac{N_1}{N} = \frac{1}{N}\sum_i D_i \to_p E[D_i]$。
- したがって、連続写像定理を適用すると：

$$
\begin{aligned}
\sqrt{N}(\bar{Y}_1 - E[Y_i(1)]) &= (1/\sqrt{N_1/N}) \cdot \sqrt{N_1}(\bar{Y}_1 - E[Y_i(1)]) \\
&\to_d (1/\sqrt{E[D_i]}) \cdot \mathrm{N}(0, Var(Y_i(1))) \\
&= \mathrm{N}\left(0, \frac{1}{E[D_i]} Var(Y_i(1))\right)
\end{aligned}
$$

漸近正規性の証明

- CLT より、$\sqrt{N_1}(\bar{Y}_1 - \mu_1) \to_d N(0, \sigma_1^2)$。
- LLN より、$\frac{N_1}{N} = \frac{1}{N}\sum_i D_i \to_p E[D_i]$。
- したがって、連続写像定理を適用すると：

$$\begin{aligned}
\sqrt{N}(\bar{Y}_1 - E[Y_i(1)]) &= (1/\sqrt{N_1/N}) \cdot \sqrt{N_1}(\bar{Y}_1 - E[Y_i(1)]) \\
&\to_d (1/\sqrt{E[D_i]}) \cdot \mathrm{N}(0, Var(Y_i(1))) \\
&= \mathrm{N}\left(0, \frac{1}{E[D_i]} Var(Y_i(1))\right)
\end{aligned}$$

- $\bar{Y}_0$ についても同様の手順を踏むことで、以下が得られます：

漸近正規性の証明

- CLT より、$\sqrt{N_1}(\bar{Y}_1 - \mu_1) \rightarrow_d N(0, \sigma_1^2)$。
- LLN より、$\frac{N_1}{N} = \frac{1}{N}\sum_i D_i \rightarrow_p E[D_i]$。
- したがって、連続写像定理を適用すると：

$$\begin{aligned}\sqrt{N}(\bar{Y}_1 - E[Y_i(1)]) &= (1/\sqrt{N_1/N}) \cdot \sqrt{N_1}(\bar{Y}_1 - E[Y_i(1)]) \\ &\rightarrow_d (1/\sqrt{E[D_i]}) \cdot \mathrm{N}(0, Var(Y_i(1))) \\ &= \mathrm{N}\left(0, \frac{1}{E[D_i]} Var(Y_i(1))\right)\end{aligned}$$

- $\bar{Y}_0$ についても同様の手順を踏むことで、以下が得られます：

$$\sqrt{N}\begin{pmatrix} \bar{Y}_1 - E[Y_i(1)] \\ \bar{Y}_0 - E[Y_i(0)] \end{pmatrix} \rightarrow_d$$

漸近正規性の証明

- CLT より、$\sqrt{N_1}(\bar{Y}_1 - \mu_1) \to_d N(0, \sigma_1^2)$。
- LLN より、$\frac{N_1}{N} = \frac{1}{N}\sum_i D_i \to_p E[D_i]$。
- したがって、連続写像定理を適用すると：

$$\begin{aligned} \sqrt{N}(\bar{Y}_1 - E[Y_i(1)]) &= (1/\sqrt{N_1/N}) \cdot \sqrt{N_1}(\bar{Y}_1 - E[Y_i(1)]) \\ &\to_d (1/\sqrt{E[D_i]}) \cdot \mathrm{N}(0, Var(Y_i(1))) \\ &= \mathrm{N}\left(0, \frac{1}{E[D_i]} Var(Y_i(1))\right) \end{aligned}$$

- $\bar{Y}_0$ についても同様の手順を踏むことで、以下が得られます：

$$\sqrt{N}\begin{pmatrix} \bar{Y}_1 - E[Y_i(1)] \\ \bar{Y}_0 - E[Y_i(0)] \end{pmatrix} \to_d \mathrm{N}\left(0, \begin{pmatrix} \frac{1}{E[D_i]} Var(Y_i(1)) & 0 \\ 0 & \frac{1}{1-E[D_i]} Var(Y_i(0)) \end{pmatrix}\right)$$

実験の仮説検定（続き）

- 以下を示しました：

$$\sqrt{N}\begin{pmatrix} \bar{Y}_1 - E[Y_i(1)] \\ \bar{Y}_0 - E[Y_i(0)] \end{pmatrix} \rightarrow_d \mathrm{N}\left(0, \begin{pmatrix} \frac{1}{E[D_i]} Var(Y_i(1)) & 0 \\ 0 & \frac{1}{1-E[D_i]} Var(Y_i(0)) \end{pmatrix}\right)$$

実験の仮説検定（続き）

- 以下を示しました：

$$\sqrt{N}\begin{pmatrix} \bar{Y}_1 - E[Y_i(1)] \\ \bar{Y}_0 - E[Y_i(0)] \end{pmatrix} \to_d \mathrm{N}\left(0, \begin{pmatrix} \frac{1}{E[D_i]}Var(Y_i(1)) & 0 \\ 0 & \frac{1}{1-E[D_i]}Var(Y_i(0)) \end{pmatrix}\right)$$

- CMT を適用すると：

$$\sqrt{N}(\bar{Y}_1 - \bar{Y}_0 - E[Y_i(1) - Y_i(0)]) \to_d N(0, \sigma^2)$$

ここで $\sigma^2 = \frac{1}{E[D_i]}Var(Y_i(1)) + \frac{1}{1-E[D_i]}Var(Y_i(0))$

実験の仮説検定（続き）

- 以下を示しました：

$$\sqrt{N}\begin{pmatrix} \bar{Y}_1 - E[Y_i(1)] \\ \bar{Y}_0 - E[Y_i(0)] \end{pmatrix} \to_d \mathrm{N}\left(0, \begin{pmatrix} \frac{1}{E[D_i]} Var(Y_i(1)) & 0 \\ 0 & \frac{1}{1-E[D_i]} Var(Y_i(0)) \end{pmatrix}\right)$$

- CMT を適用すると：

$$\sqrt{N}(\bar{Y}_1 - \bar{Y}_0 - E[Y_i(1) - Y_i(0)]) \to_d N(0, \sigma^2)$$

ここで $\sigma^2 = \frac{1}{E[D_i]} Var(Y_i(1)) + \frac{1}{1-E[D_i]} Var(Y_i(0))$

- これにより、$\tau = E[Y_i(1) - Y_i(0)]$ に対する 95% 信頼区間を形成できます：

$$\bar{Y}_1 - \bar{Y}_0 \pm 1.96\hat{\sigma}/\sqrt{N}$$

ここで $\hat{\sigma}^2 = \frac{N}{N_1}\hat{\sigma}_1^2 + \frac{N}{N_0}\hat{\sigma}_0^2$ であり、$\hat{\sigma}_d^2$ はグループ $d \in \{0,1\}$ の標本分散です。

うつ病の結果に関する標本平均（再掲）

	対照群	処置群
平均	0.329	0.306
標準偏差	0.470	0.461
サンプルサイズ (N)	10426	13315

うつ病の結果に関する標本平均（再掲）

	対照群	処置群
平均	0.329	0.306
標準偏差	0.470	0.461
サンプルサイズ (N)	10426	13315

- 処置効果の点推定値は $\hat{\tau} =$

うつ病の結果に関する標本平均（再掲）

	対照群	処置群
平均	0.329	0.306
標準偏差	0.470	0.461
サンプルサイズ (N)	10426	13315

- 処置効果の点推定値は $\hat{\tau} = 0.306 - 0.329 = -0.023$ です。

うつ病の結果に関する標本平均（再掲）

	対照群	処置群
平均	0.329	0.306
標準偏差	0.470	0.461
サンプルサイズ (N)	10426	13315

- 処置効果の点推定値は $\hat{\tau} = 0.306 - 0.329 = -0.023$ です。
- 処置効果の信頼区間（CI）は：

うつ病の結果に関する標本平均（再掲）

	対照群	処置群
平均	0.329	0.306
標準偏差	0.470	0.461
サンプルサイズ (N)	10426	13315

- 処置効果の点推定値は $\hat{\tau} = 0.306 - 0.329 = -0.023$ です。
- 処置効果の信頼区間（CI）は：

$$\hat{\tau} \pm 1.96 \times \sqrt{\frac{1}{N_1}\hat{\sigma}_1^2 + \frac{1}{N_0}\hat{\sigma}_0^2} =$$

うつ病の結果に関する標本平均（再掲）

	対照群	処置群
平均	0.329	0.306
標準偏差	0.470	0.461
サンプルサイズ (N)	10426	13315

- 処置効果の点推定値は $\hat{\tau} = 0.306 - 0.329 = -0.023$ です。
- 処置効果の信頼区間（CI）は：

$$\hat{\tau} \pm 1.96 \times \sqrt{\frac{1}{N_1}\hat{\sigma}_1^2 + \frac{1}{N_0}\hat{\sigma}_0^2} =$$

$$-0.023 \pm 1.96 \times \sqrt{\frac{1}{13315}0.461^2 + \frac{1}{10426}0.470^2}$$

$$=$$

うつ病の結果に関する標本平均（再掲）

	対照群	処置群
平均	0.329	0.306
標準偏差	0.470	0.461
サンプルサイズ (N)	10426	13315

- 処置効果の点推定値は $\hat{\tau} = 0.306 - 0.329 = -0.023$ です。
- 処置効果の信頼区間（CI）は：

$$
\begin{aligned}
&\hat{\tau} \pm 1.96 \times \sqrt{\frac{1}{N_1}\hat{\sigma}_1^2 + \frac{1}{N_0}\hat{\sigma}_0^2} = \\
&-0.023 \pm 1.96 \times \sqrt{\frac{1}{13315}0.461^2 + \frac{1}{10426}0.470^2} \\
&= [-0.035, -0.001]
\end{aligned}
$$

非交絡性の下での仮説検定

- 非交絡性 $D_i \perp\!\!\!\perp (Y_i(1), Y_i(0)) | X_i$ の下では：

$$\underbrace{E[Y_i(1) - Y_i(0) | X_i = x]}_{CATE(x)} = E[Y_i | D_i = 1, X_i = x] - E[Y_i | D_i = 0, X_i = x]$$

つまり、各 X_i の値において実験が行われているのと同じです。

非交絡性の下での仮説検定

- 非交絡性 $D_i \perp\!\!\!\perp (Y_i(1), Y_i(0))|X_i$ の下では：

$$\underbrace{E[Y_i(1) - Y_i(0)|X_i = x]}_{CATE(x)} = E[Y_i|D_i = 1, X_i = x] - E[Y_i|D_i = 0, X_i = x]$$

  つまり、各 X_i の値において実験が行われているのと同じです。
- 実験の場合と同じ論理により：

$$\sqrt{N_x}(\bar{Y}_{1,x} - \bar{Y}_{0,x} - CATE(x)) \rightarrow_d N(0, \sigma_x^2)$$

  ここで $N_x = |i : X_i = x|$ および $\sigma_x^2 = \frac{Var(Y_i(1)|X_i=x)}{E[D_i|X_i=x]} + \frac{Var(Y_i(0)|X_i=x)}{E[1-D_i|X_i=x]}$ です。

非交絡性の下での仮説検定

- 非交絡性 $D_i \perp\!\!\!\perp (Y_i(1), Y_i(0))|X_i$ の下では：

$$\underbrace{E[Y_i(1) - Y_i(0)|X_i = x]}_{CATE(x)} = E[Y_i|D_i = 1, X_i = x] - E[Y_i|D_i = 0, X_i = x]$$

  つまり、各 X_i の値において実験が行われているのと同じです。

- 実験の場合と同じ論理により：

$$\sqrt{N_x}(\bar{Y}_{1,x} - \bar{Y}_{0,x} - CATE(x)) \to_d N(0, \sigma_x^2)$$

  ここで $N_x = |i : X_i = x|$ および $\sigma_x^2 = \frac{Var(Y_i(1)|X_i=x)}{E[D_i|X_i=x]} + \frac{Var(Y_i(0)|X_i=x)}{E[1-D_i|X_i=x]}$ です。

- したがって、N_x が大きい場合には $CATE(x)$ についても仮説検定を行うことができます。

非交絡性の下での仮説検定

- 非交絡性 $D_i \perp\!\!\!\perp (Y_i(1), Y_i(0))|X_i$ の下では：

$$\underbrace{E[Y_i(1) - Y_i(0)|X_i = x]}_{CATE(x)} = E[Y_i|D_i = 1, X_i = x] - E[Y_i|D_i = 0, X_i = x]$$

  つまり、各 X_i の値において実験が行われているのと同じです。
- 実験の場合と同じ論理により：

$$\sqrt{N_x}(\bar{Y}_{1,x} - \bar{Y}_{0,x} - CATE(x)) \to_d N(0, \sigma_x^2)$$

  ここで $N_x = |i : X_i = x|$ および $\sigma_x^2 = \frac{Var(Y_i(1)|X_i=x)}{E[D_i|X_i=x]} + \frac{Var(Y_i(0)|X_i=x)}{E[1-D_i|X_i=x]}$ です。
- したがって、N_x が大きい場合には $CATE(x)$ についても仮説検定を行うことができます。
- $CATE(x)$ を平均化することで、ATE に関する推論も可能です。

連続的な x の課題

- これまでは、$X_i = x$ である観測数が多い場合に $CATE(x)$ を推定する方法を示してきました。
- これは X_i が 2 値（例：大学卒かどうか）であったり、少数の離散値（例：50 の州）であったりする場合には非常にうまく機能します。

連続的な x の課題

- これまでは、$X_i = x$ である観測数が多い場合に $CATE(x)$ を推定する方法を示してきました。
- これは X_i が 2 値（例：大学卒かどうか）であったり、少数の離散値（例：50 の州）であったりする場合には非常にうまく機能します。
- しかし、X_i が連続量である場合はどうでしょうか？

連続的な x の課題

- これまでは、$X_i = x$ である観測数が多い場合に $CATE(x)$ を推定する方法を示してきました。
- これは X_i が 2 値（例：大学卒かどうか）であったり、少数の離散値（例：50 の州）であったりする場合には非常にうまく機能します。
- しかし、X_i が連続量である場合はどうでしょうか？
- 例えば X_i が所得である場合、$CATE(50, 351)$ を推定するには、理論上、所得が正確に 50,351 ドルである処置群と対照群の観測値が大量に必要です。ほとんどのデータセットでは、正確にこの所得を持つ人はそれほど多くありません。

連続的な x の課題

- これまでは、$X_i = x$ である観測数が多い場合に $CATE(x)$ を推定する方法を示してきました。
- これは X_i が 2 値（例：大学卒かどうか）であったり、少数の離散値（例：50 の州）であったりする場合には非常にうまく機能します。
- しかし、X_i が連続量である場合はどうでしょうか？
- 例えば X_i が所得である場合、$CATE(50,351)$ を推定するには、理論上、所得が正確に 50,351 ドルである処置群と対照群の観測値が大量に必要です。ほとんどのデータセットでは、正確にこの所得を持つ人はそれほど多くありません。
- したがって、X_i が連続的に分布している場合には、条件付き期待値を推定する別の方法が必要です。

連続的な x の課題

- これまでは、$X_i = x$ である観測数が多い場合に $CATE(x)$ を推定する方法を示してきました。
- これは X_i が 2 値（例：大学卒かどうか）であったり、少数の離散値（例：50 の州）であったりする場合には非常にうまく機能します。
- しかし、X_i が連続量である場合はどうでしょうか？
- 例えば X_i が所得である場合、$CATE(50,351)$ を推定するには、理論上、所得が正確に 50,351 ドルである処置群と対照群の観測値が大量に必要です。ほとんどのデータセットでは、正確にこの所得を持つ人はそれほど多くありません。
- したがって、X_i が連続的に分布している場合には、条件付き期待値を推定する別の方法が必要です。
- コースの次のパートでは、CEF の近似として線形回帰を用いることで、このタスクを達成する方法に焦点を当てます。