統計学入門
～推定量とその性質～

2022年度1学期：　　月曜2限

担当教員：　　　　　石垣　司

推定

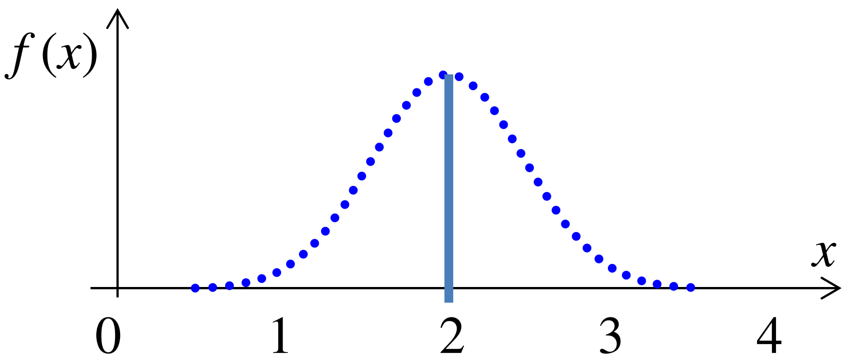
標本から母集団の分布のパラメータ(母数)を求める

- **点推定**: 未知のパラメータ θ を1点で推定
- **区間推定**: 未知のパラメータ θ が存在する区間を推定

- データ(標本の実現値)を利用して未知 θ の値を推定

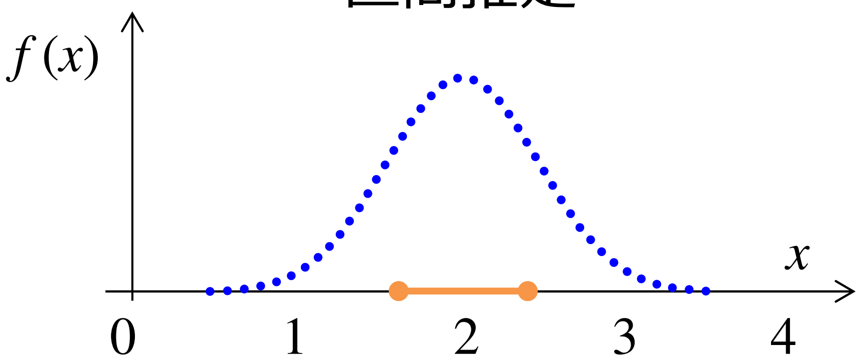
  例 / データ$\{2, 3, 1, 2, 3, 4, \ldots, 2\}$から母集団の母平均を推定

点推定



$\hat{\theta} = 2$

区間推定



$P(1.7 \leq \hat{\theta} \leq 2.3)$

推定量 #1

推定量(estimator)

- 標本からパラメータを推定する関数

  統計量と同じ。推定が目的の場合の統計量。推定量も確率変数

  推定値：観測値から推定量を計算した実際の値
- 無作為標本 $\boldsymbol{X} = \{X_1, \dots, X_n\}$
- パラメータ θ
- 推定量 $T(\boldsymbol{X})$ or $\hat{\theta}$

- 例：標本平均の推定量

$$\hat{\theta} = T(\boldsymbol{X}) = \bar{X} = \underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i}_{\text{推定量}}$$

推定量 #2

推定量は自由に設定可能

– **例: 母平均 μ の推定量の候補**

$$T_1(\boldsymbol{X}) = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$T_2(\boldsymbol{X}) = \frac{1}{n+1}\left(X_1 + \sum_{i=1}^{n} X_i\right)$$

$$T_3(\boldsymbol{X}) = \frac{1}{2}(X_1 + X_n)$$

$T_1(X)$ が標本平均の推定量として用いられるのは, 不偏性, 一致性, よい効率性をもつため

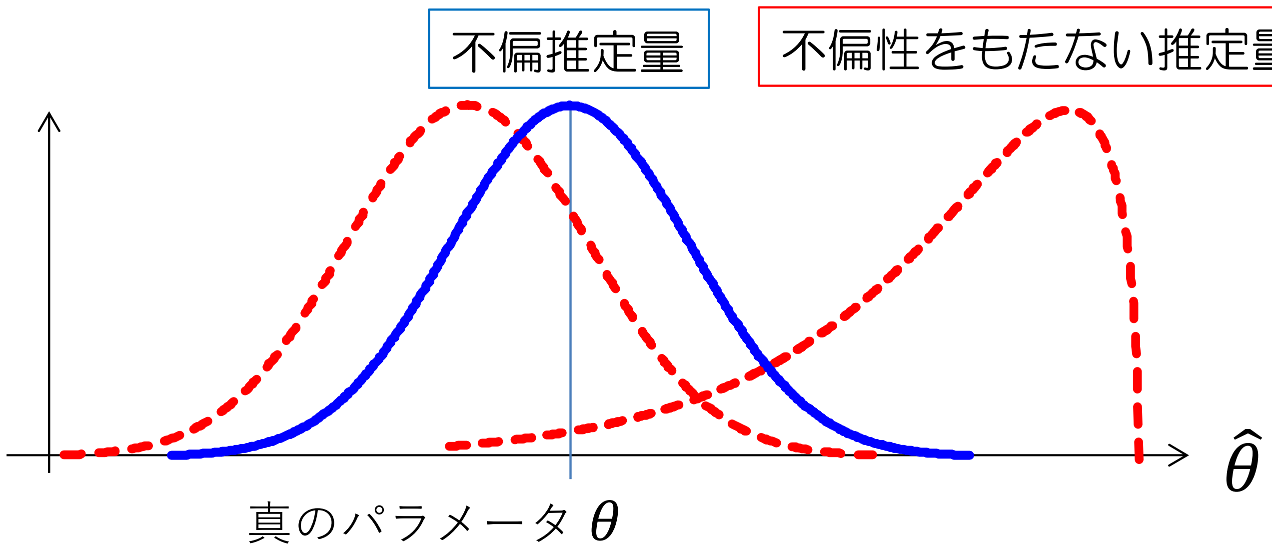
不偏性(Unbiasedness)

推定量の期待値がパラメータ θ となる性質

$$E[T(\boldsymbol{X})] = \theta$$

不偏推定量：不偏性をもつ推定量

- 「推定量 $T(X)$ は不偏である」
- 平均的に過大評価も過小評価もしない推定量



演習問題

問題

- **推定量 $T_1(\boldsymbol{X})$ と $T_2(\boldsymbol{X})$ はそれぞれ母平均の不偏推定量となるか確かめよ**

$$T_1(\boldsymbol{X}) = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$T_2(\boldsymbol{X}) = \frac{1}{n+1}\left(X_1 + \sum_{i=1}^{n} X_i\right)$$

- **次の推定量 $\tilde{S}^2$ は母分散の不偏推定量となるか確かめよ**

$$\tilde{S}^2 = \frac{1}{n}\sum_{i=1}^{n} (X_i - \bar{X})^2$$

一致性(Consistency)

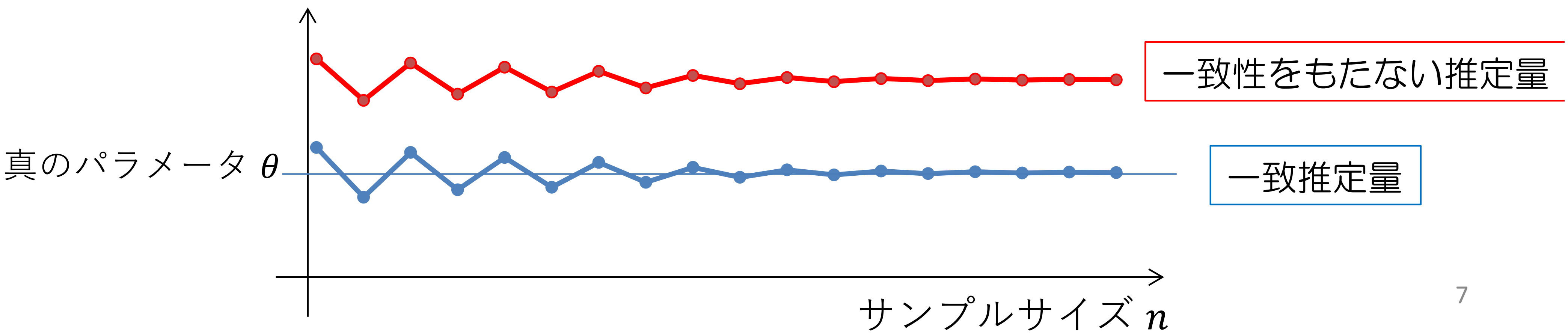
サンプルサイズ n が大きくなると推定量がパラメータ θに近づく性質

$$\lim_{n\to\infty} P(|T(\boldsymbol{X}) - \theta| > \varepsilon) = 0 \quad (\varepsilon \text{ は正の定数})$$

推定量 $T(\boldsymbol{X})$ と母数 θ のズレが ε よりも大きい確率はゼロに近づく

一致推定量：一致性をもつ推定量

– **推定量が一致性を持たない場合，サンプルサイズ n をどれだけ大きくしても，母集団のパラメータ(一点)に一致しない**



一致性の判定

一致性を満たす推定量の例

check!

- 母平均の推定量　標本平均 $T_1(\boldsymbol{X}) = \bar{X}, T_2(\boldsymbol{X})$
- 母分散の推定量　標本分散 $S^2, \tilde{S}^2$

一致性を満たさない推定量 $T(\boldsymbol{X})$ のチェック方法

- $\lim_{n\to\infty} V[T(\boldsymbol{X})] > 0$ ならば一致推定量とはならない

  $\lim_{n\to\infty} V[T(\boldsymbol{X})] = 0$ は一致推定量の必要条件

  十分条件や必要十分条件ではないので注意

問題

- $T_3(\boldsymbol{X}) = \frac{1}{2}(X_1 + X_n)$は母平均の不偏推定量となるか，また，一致推定量となるかそれぞれ確かめよ

標本平均と標本分散の統計量の性質

標本分散の候補となる統計量の性質

- $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$　不偏性○　一致性○
- $\tilde{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$　不偏性×　一致性○
- **S^2 の方が $\tilde{S}^2$ よりも統計量として優れている**

標本平均の候補となる統計量の性質

- $T_1(\boldsymbol{X}) = \frac{1}{n}\sum_{i=1}^{n} X_i$　不偏性○　一致性○
- $T_2(\boldsymbol{X}) = \frac{1}{n+1}\left(X_1 + \sum_{i=1}^{n} X_i\right)$　不偏性○　一致性○
- $T_3(\boldsymbol{X}) = \frac{1}{2}(X_1 + X_n)$　不偏性○　一致性×

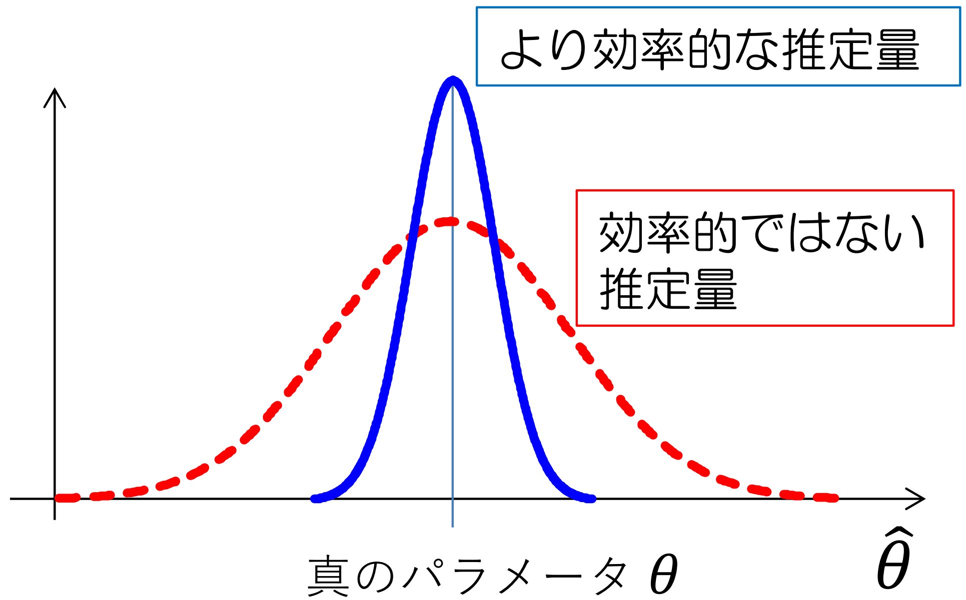
なぜ標本平均の統計量として T_1 が用いられるのか？

効率性(Efficiency)

効率的な推定量：分散のより小さな推定量

最小分散推定量：最も分散の小さな推定量

不偏推定量かつ一致推定量が複数ある場合，さらに良い推定量を選ぶ基準として用いられる



問題

$$T_1(\boldsymbol{X}) = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$T_2(\boldsymbol{X}) = \frac{1}{n+1}\left(X_1 + \sum_{i=1}^{n} X_i\right)$$

– $T_1(X)$と $T_2(X)$ **は不偏推定量かつ一致推定量である。$V[T_1(X)]$ と $V[T_2(X)]$ からどちらがより効率的な推定量となるか確かめよ**

その他の性質

最小分散不偏推定量

– あらゆる不偏推定量の中で, 最も効率的な推定量

正規母集団 $N(\mu, \sigma^2)$ からの標本に関して,

「標本平均 $\bar{X}$ は母平均 μ の最小分散不偏推定量」

「標本分散 S^2 は母分散 σ^2 の最小分散不偏推定量」

"平均"や"分散"を当然のように推定量として使用しているが,
その理論的正当性の保証

その他の性質

– 漸近正規性, 漸近有効性, 十分統計量, 頑健性など

補足：なぜこのような規範が必要？

不偏性や一致性の必要性

- 平均や分散の普遍性や一致性は自明に近い
- 自明ではない例：単回帰分析の最小2乗推定量

$$\hat{b} = \frac{\sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y}}{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}, \hat{a} = \bar{Y} - \hat{b}\bar{X}$$

- $\hat{b}$ は母集団の傾きの不偏推定量？一致推定量？

  ちなみに, ある仮定の下で, 最小分散不偏推定量かつ一致推定量

経済データ分析の特徴

- 現象は複雑だが観測できるサンプルサイズ n は小さい
- 不偏性や一致性を持つ推定量での議論が重要