講義にて配布した資料の抜粋・縮小版です



# 統計工学　講義資料(5)

～統計的推論：推定～

(C) Sadami SUZUKI



## 2つのタイプの統計的推論

### 推定

- 標本に基づいて母集団の性質は○○であると，その値を求めること

**点推定**・・・標本から母集団の性質について，一つの値を求めること

**区間推定**・・・2つの統計値を求め，その中に入っていると思われる区間を求めること

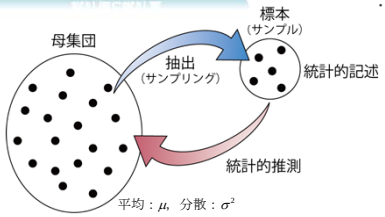
### 検定

- 標本について観察・調査した結果からYesまたはNoなる判定を下すこと
- 経験等に基づき，母集団のある性質について，その値は○○であるとの仮定(仮説)をたて，それが成立しているかを標本に照らして判断すること

(C) Sadami SUZUKI



## 統計値と統計量




- サンプルを**母集団分布**に従う**確率変数**として捉え，その関数**統計量**を考える

観測値：$x_1, x_2, \cdots, x_n$

確率変数：$X_1, X_2, \cdots, X_n$

- **統計量**・・・サンプルに対応する**確率変数**の関数：$g(X)=\sum_{i=1}^{n} X_i$
- **統計値**・・・統計量の実現値

(C) Sadami SUZUKI



## 母集団のパラメータ(母数)を推定する

- 確率変数 $X_1, X_2, \cdots, X_n$ の実現値とみなされるサンプル $x_1, x_2, \cdots, x_n$ から，母集団 $X=(X_1, X_2, \cdots, X_n)$ の確率分布のパラメータ(母数)の値を推測することを**推定**といい，推定に用いられる統計量のことを**推定量**という

### 推定量

- 母数を推定するための**ルール**

（例）$g(X)=\bar{X}=\dfrac{\sum X_i}{n}$

### 推定値

- 標本観察によって実際に得られた値

（例）$\bar{x}=50.5$

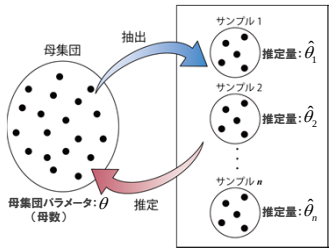
| 番号 | $x_i$ |
|---|---|
| 1 | 50.8 |
| 2 | 50.3 |
| 3 | 49.6 |
| 4 | 50.6 |
| 5 | 51.5 |
| 6 | 51.0 |
| 7 | 50.8 |
| 8 | 49.2 |
| 9 | 50.7 |

(C) Sadami SUZUKI



## 推定のためのルールを評価しよう

- **推定量**は標本からルール(関数)に基づいて計算されることから，標本が異なればその値も異なる
- **確率変数**としての**推定量**の変動(バラツキ)は不可避



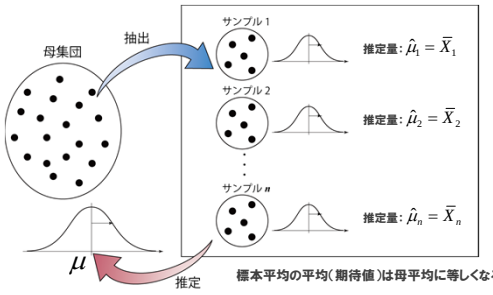

**推定量の不偏性**

(C) Sadami SUZUKI



## 母平均の推定量

- **標本平均**は母平均の**不偏推定量**




標本平均の平均(期待値)は母平均に等しくなる

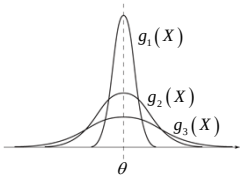
(C) Sadami SUZUKI

# 講義にて配布した資料の抜粋・縮小版です



## 推定のためのルールを評価しよう

- 母集団パラメータは**不偏推定量**によって**平均的**に正しく推定することが可能
- 分散が最小である不偏推定量を**最良不偏推定量**という




**クラーメル・ラオの不等式**

$X_1,\cdots,X_n$ が互いに独立に確率密度関数 $f(x;\theta)$ に従い，
$\hat{\theta}(X)$ を $\theta$ に対する任意の不偏推定量とするとき，
推定量の分散について

$$V\left[\hat{\theta}(X)\right] \geq \frac{1}{nI_x(\theta)}$$

ここに，$I_x(\theta)=E\left[S^2(X;\theta)\right]$，$S(X;\theta)=\frac{d}{d\theta}\log f(x;\theta)$

の不等式が成り立つ．右辺は分散の下限を与える．
とくに等号が成り立つ不偏推定量を有効推定量という．
（※有効推定量が存在すれば，それは最良不偏推定量である）





## 推定のためのルールを評価しよう

確率変数：$X_1, X_2, \cdots, X_n$ による母平均 $\mu$ の推定

$$E[X_i]=\mu,\ V[X_i]=\sigma^2,\ Cov[X_i, X_j]=0\ \ (i \neq j)$$

$X_1, X_2, \cdots, X_n$ の任意の 1次結合 $g(X)=a_1X_1+a_2X_2+\cdots+a_nX_n$
が $\mu$ の不偏推定量であるためには

$$E[g(X)]=\mu$$

が成り立たなければならない．

$$E[g(X)]=a_1\mu+a_2\mu+\cdots+a_n\mu=(a_1+a_2+\cdots+a_n)\mu$$

であり，$\mu$ は任意であることから，$a_1+a_2+\cdots+a_n=1\cdots(*)$ が条件となる．一方，

$$V[g(X)]=a_1^2\sigma^2+a_2^2\sigma^2+\cdots+a_n^2\sigma^2=(a_1^2+a_2^2+\cdots+a_n^2)\sigma^2$$

において，$a_1^2+a_2^2+\cdots+a_n^2$ が条件 (*) の下で最少となるのは

$$a_1=a_2=\cdots=a_n=\frac{1}{n}$$

のときである．

- 標本平均 $\bar{X}$ は母平均 $\mu$ の**最良不偏推定量**である

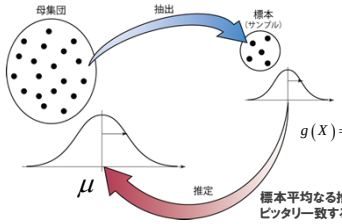




## 母平均の推定

- 抽出したサンプルを**確率変数**と考えると・・・

観測値：$x_1, x_2, \cdots, x_n$ → 確率変数：$X_1, X_2, \cdots, X_n$
による母平均 $\mu$ の推定




$$g(X)=\bar{X}=\frac{\sum X_i}{n}$$

標本平均なる推定値は必ずしも母平均にピッタリ一致するわけではない

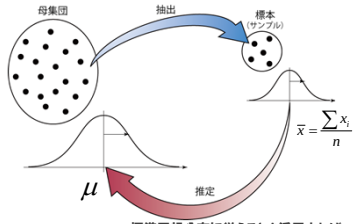
標本平均から，**母平均が含まれる範囲（区間）**を推定しよう





## 母平均の区間推定

- 母平均は**「この値からこの値までの間に含まれるのではないか」**という形で推定しよう
- 母集団や標本平均の分布に着目すると・・・




$$\bar{x}=\frac{\sum x_i}{n}$$

- 母集団が正規分布に従うとき
- サンプル数が大きいとき（近似的に）

$$z=\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$$

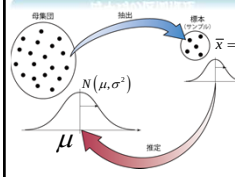
**標準正規分布に従う**

標準正規分布に従うことを活用すれば母平均が含まれる区間を求めることができるはず

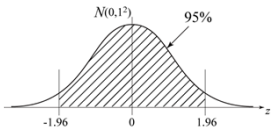




## 母平均の区間推定




$$\bar{x}=\frac{\sum x_i}{n} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad \rightarrow \quad z=\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,1^2)$$




$N(0,1)$ に従う変数が $-1.96$から$+1.96$ の間の値をとる確率が95%であることを利用して…

$$\Pr\left\{-1.96 < \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} < 1.96\right\}=0.95$$

$$\Pr\left\{\bar{x}-1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{x}+1.96\frac{\sigma}{\sqrt{n}}\right\}=0.95$$

**平均値 $\mu$ の95%信頼区間**





**【確認用演習課題】**

1. プロ野球のA選手の打撃成績は30打数9安打であり，B選手の打撃成績は340打数102安打であった．
   ともに打率は3割であるが，両者の違いを統計的推論の立場から考察せよ．

2. 分散$\sigma^2=36$ の母集団から $n=100$ の標本を取り出したところ，$\bar{x}=50.5$ であった．
   このとき，母集団平均を信頼区間 90%, 95%, 99% で区間推定せよ．