

確率と統計

中山クラス 第8週

0

4.5 標本分布

標本分布とは

標本統計量(標本平均, 標本分散など)に関する確率分布
確率変数である標本統計量の分布状況を確率的に表す
→母数の推定値の信頼度を知る上で重要

標本における個々のデータの実現値を表した度数分布ではなく、
標本統計量の確率分布である。

母集団分布と標本統計量の種類, サンプルサイズから理論的
(数学的)に導かれるもので, 実際のデータから作成されるものではない。

3

本日の内容

小テストの解説

第4章

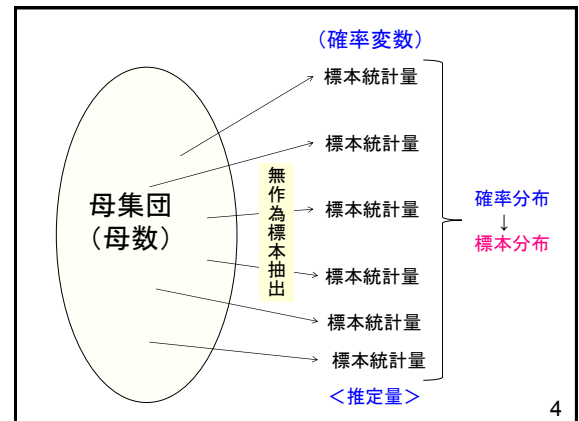
4.5 標本分布

4.6 標本平均以外の標本分布

第3回レポートの出題

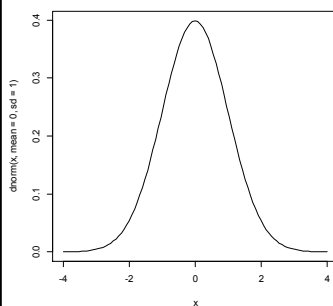
コンピュータ演習

1



4

正規分布に関する補足



$$P(|X - \mu| < \sigma) = 0.6826$$

$$P(|X - \mu| < 2\sigma) = 0.9546$$

$$P(|X - \mu| < 3\sigma) = 0.9974$$

2

4.5.1 標本分布から分かること

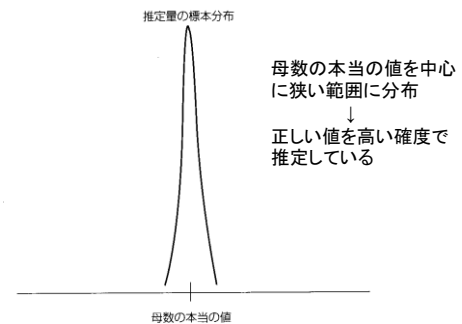
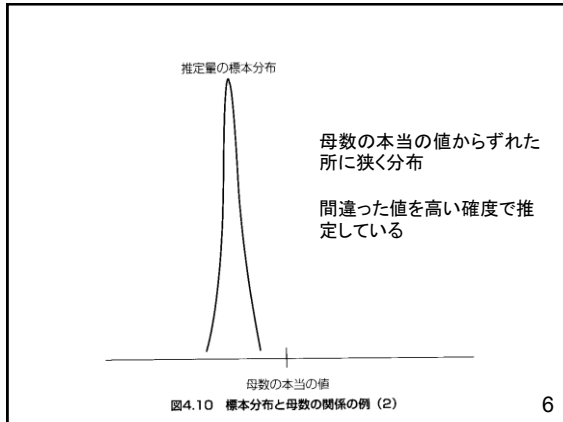


図4.9 標本分布と母数の関係の例 (1)

5

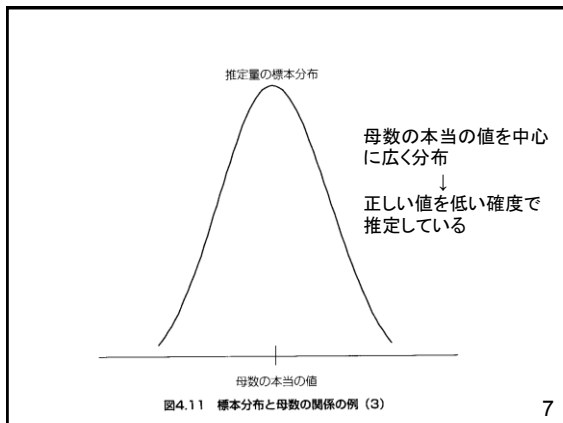


6

4.5.2 標本分布を「経験的」に求める

確率変数の実現値を多量に得る
↓
実現値のヒストグラムが確率分布に近づく
↓
近似的な標本分布

9



7

4.5.3 正規母集団の母平均の推定

正規分布に従う母集団の平均(=母平均)を推定する。

→標本の平均(=標本平均)=母平均の推定量



Rで標本平均を計算してみると、試行ごとに標本平均が変化する。

(Rでは平均をあらかじめ決めてサンプルを生成しているので、真の母平均がわかっている。教科書の例では50が母平均。)



標本抽出を繰り返して推定値を調べる。→4.5.4

10

「標本分布の平均と標準偏差(標準誤差)」により、推定値がどれだけ母数に近いか評価できる。

標本分布は、理論的に計算できるが、簡単ではない。
→Rを使って「経験的(実験的)に」標本分布を計算

注意:

- サンプルから計算する標本分布は真の標本分布の「近似値」にすぎない。
- 母集団の分布は種々あるが本章では正規分布を想定する。

8

母集団分布: $N(50, 10^2)$
サンプル数: $n=10$

標本の無作為抽出と平均

```
> 標本 <- rnorm(n=10, mean=50, sd=10)
```

```
> 標本
```

```
[1] 49.78527 53.83358 32.68944 50.65051 55.88846 54.73556  
40.45287 52.80188  
[9] 65.50508 62.29281
```

```
> mean(標本)
```

```
[1] 51.86355
```

11

```
> 標本 <- rnorm(n=10, mean=50, sd=10)
> 標本
[1] 49.78527 53.83358 32.68944 50.65051 55.88846 54.73556
40.45287 52.80188
[9] 65.50508 62.29281
> mean(標本)
[1] 51.86355

> 標本 <- rnorm(n=10, mean=50, sd=10)
> 標本
[1] 56.55998 62.13028 63.16473 52.05961 61.03372 57.48173
50.13165 45.09212
[9] 51.82399 37.71204
> mean(標本)
[1] 53.71898
```

12

標本平均の確率分布

正規分布 $N(\mu, \sigma^2)$ に従う母集団から n サンプルを無作為抽出したとき、その標本平均の確率分布 (標本分布) は $N(\mu, \frac{\sigma^2}{n})$ に従う。

```
> mean(標本平均)
[1] 50.00082
```

```
> mean((標本平均-mean(標本平均))^2)
[1] 9.906092
```

15

4.5.4 標本分布を求める

<推定値を格納する場所を確保>

```
> 標本平均 <- numeric(length=10000)
```

<{}で囲まれた処理を10000回繰り返す>

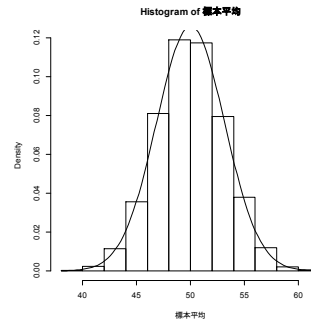
```
> for(i in 1:10000){
+   標本<- rnorm(n=10,mean=50,sd=10)
+   標本平均[i]<- mean(標本) }
```

<ヒストグラムを表示>

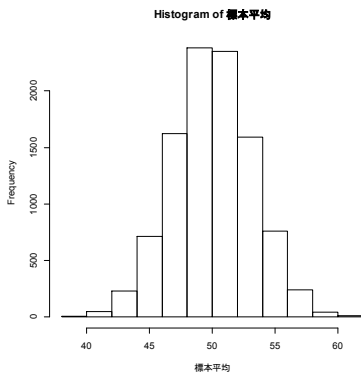
```
> hist(標本平均)
```

13

```
> hist(標本平均,freq=FALSE) <面積=1となるように調整>
> curve(dnorm(x,mean=50,sd=sqrt(10)),add=TRUE)
```



16



14

4.5.5 不偏性

ある推定量の標本分布の平均が推定しようとしている母数の値と一致するとき、その推定量は**不偏性がある** (不偏である)。→ **不偏推定量**

標本平均は母集団分布にかかわらず母平均の不偏推定量である。

推定量の不偏性→標本分布が母数の**本当の値**を中心として分布している。

17

4.5.6 標準誤差

標準誤差: 標本分布の広がり(標準偏差で評価)

$N(50, 10^2)$ の正規分布から $n = 10$ の標本を抽出したときの
標本平均の標本分布は $N(50, 10)$ であったから、標準誤差は $\sqrt{10}$
となる。

$N(\mu, \sigma^2)$ の母集団に対して標本分散は $N\left(\mu, \frac{\sigma^2}{n}\right)$ となるので、
標準誤差は $\sigma/\sqrt{n}$ となる。

母集団分布の分散(標準偏差)が大きい
→ 標本平均の標準誤差も大きい
サンプルサイズ n が大きい
→ 標本平均の標準誤差は小さい

18

標本分散と不偏分散の標準偏差

```
> sd(標本分散)
```

```
[1] 42.61138
```

```
> sd(不偏分散)
```

```
[1] 47.34598
```

不偏分散の方がばらつきが大きい

21

4.6 標本平均以外の標本分布

平均以外の母数に対しても標本分布を考えて母
数を推定できる。

4.6.1 標本分散と不偏分散の標本分布

標本分散: 標本データに基づく分散(分母:n)

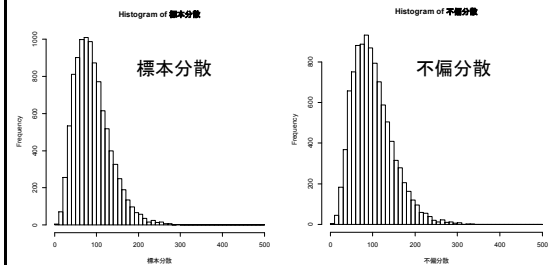
不偏分散: 母分散の不偏推定量(分母:n-1)

実験的に違いを調べる

19

```
> hist(標本分散,breaks=seq(0,500,10))
```

```
> hist(不偏分散,breaks=seq(0,500,10))
```



22

母集団分布: $N(50, 10^2)$, サンプルサイズ: $n = 10$
母分散=100

```
> 標本分散 <- numeric(length=10000)
```

```
> 不偏分散 <- numeric(length=10000)
```

```
> for(i in 1:10000){
+   標本<- rnorm(n=10,mean=50,sd=10)
+   標本分散[i]<-mean((標本-mean(標本))^2)
+   不偏分散[i]<-var(標本)
+ }
```

```
> mean(標本分散)    <標本分散は1/n>
```

```
> [1] 90.46028
```

```
> mean(不偏分散)    <不偏分散は1/(n-1)>
```

```
> [1] 100.5114
```

20

不偏分散の標準偏差

```
> mean(sqrt(不偏分散))
```

```
[1] 9.752271
```

母標準偏差=10 > 9.75

不偏分散の平方根は母標準偏差の推定量にはならない。

23

4.6.2 中央値の標本分布

中央値: データを並べて中央に位置する値. 平均よりも外れ値の影響を受けにくい.

→ 標本中央値の標本分布を実験的に求める.

→ 標本中央値の推定値の平均は母平均に近い. しかし, 標準誤差は標本平均よりも大きくなる. (実は, 標本中央値のほうが外側の値が混ざる→図4.16)

→ 母平均の推定量としては, 標本中央値よりも標本平均のほうが平均的に誤差が小さく, 適している.

24

第3回レポート

◆ 締切: 12月6日(金) 17:00

◆ 提出場所: 1号館2階 レポート提出箱

27

中央値の標本分布

```
> 標本平均 <- numeric(length=10000)
> 標本中央値 <- numeric(length=10000)
> for(i in 1:10000){
+ 標本<- rnorm(n=10, mean=50, sd=10)
+ 標本平均[i]<-mean(標本)
+ 標本中央値[i]<-median(標本)
+ }
> mean(標本平均)
[1] 49.96765
> mean(標本中央値)
[1] 49.98527
> sd(標本平均)
[1] 3.160037
> sd(標本中央値)
[1] 3.75616
```

25

第3回レポート課題

I. 次の用語を説明せよ.

母集団, 母数, 標本, 標本抽出, 推定量, 推定値, 確率変数, 確率分布, 正規分布, 標本分布, 不偏性, 標準誤差

II. 第4章の練習問題を行ったうえで, 以下に答えよ.

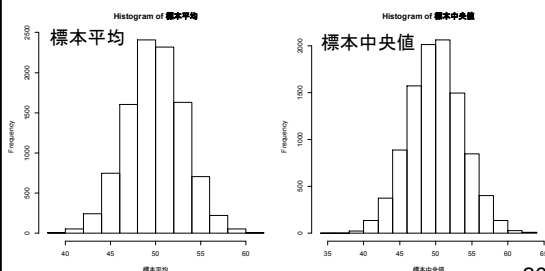
(1) で作成した図から分かることを述べよ.

(2) で作成した図から分かることを述べよ.

* (1), (2) に関して, 図のプリントアウトに書き込んでもよい.

28

```
> hist(標本平均)
> hist(標本中央値)
```



26

次回の予定

第5章 統計的仮説検定

5. 1 ~ 5. 3

標準正規分布を用いた検定を行う予定

29