

# データ解析のための統計学入門

## 第5章

### GLMの尤度比検定と 検定の非対称性

大内 克之

# 検定

検定は、推定された統計モデルを比較する方法のひとつ。

どんな統計モデルにでも利用できる検定に**尤度比検定**がある。

# 検定とAIC

統計モデル検定

AICによるモデル選択

解析対象のデータを確定



データを説明できるような統計モデルを設定

(帰無仮説・対立仮説)



(単純モデル・複雑モデル)

ネストした統計モデルたちのパラメータの最尤推定計算



帰無仮説棄却の危険率を評価



モデル選択基準AICの評価



帰無仮説棄却の可否を判断



予測の良いモデルを選ぶ

# 例題

・一定モデル: 種子数の平均 $\lambda_i$ が定数であり、体サイズ $x_i$ に依存しないモデル

・xモデル: 種子数の平均 $\lambda_i$ が体サイズ $x_i$ に依存するモデル

---

モデル	k	logL*	deviance -2logL*	residual deciance	AIC
一定	1	-237.6	475.3	89.5	477.3
x	2	-235.4	470.8	85.0	474.8
フル	100	-192.9	385.8	0.0	585.8

---

# 尤度比と逸脱度

尤度比

$$\frac{L_1^*}{L_2^*} = \frac{\text{一定モデルの最大尤度: } \exp(-237.6)}{x \text{ モデルの最大尤度: } \exp(-235.4)}$$

逸脱度

$$D_{1,2} = -2(\log L_1^* - \log L_2^*)$$

例題データの逸脱度は $D_{1,2}=4.5$ 程度。

# 検定における2種類の過誤

比較するモデルを帰無仮説と対立仮説に分類する。

- ・帰無仮説：一定モデル
- ・対立仮説：xモデル

---

観察された逸脱度  $\Delta D_{1,2}$  は

↓ 帰無仮説は

「めったにない差」  
(帰無仮説を棄却)

「よくある差」  
(棄却できない)

---

真のモデルである

第一種の過誤

(問題なし)

---

真のモデルではない

(問題なし)

第二種の過誤

---

# 尤度比と逸脱度

P値の大小を判断するため、有意水準 $\alpha$ を自分で決める。 $\alpha=0.05$ などがよく使われる。

- ・  $P \geq \alpha \rightarrow$  帰無仮説棄却できない。
- ・  $P < \alpha \rightarrow$  帰無仮説棄却できる。

# P値の計算方法

P値の計算には、

- ・パラメトリックブートストラップ (PB) 法
  - ・ $\chi^2$ を使った近似計算法
- などがある。

# PB法

PB法の手順は、

- (1)平均種子数と同じ平均の乱数を発生させる。
- (2)それに対して、一定モデル、xモデルのglm()の推定。
- (3)逸脱度の差の計算。

# PB法の関数

<http://hosho.ees.hokudai.ac.jp/~kubo/ce/IwanamiBook.html>

上のサイトの第五章にあるdata3a.csvと、pb.Rをダウンロードし、Rの作業ディレクトリに入れる。

関数の実行前に、data3a.csvとpb.Rを読み込んでおく。

```
> d <- read.csv("data3a.csv")  
> source("pb.R")  
> dd12 <- pb(d, n.bootstrap = 1000) }
```

# 関数の実行

summaryを使って概要を見る。

```
> summary(dd12)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.000009	0.114100	0.455200	0.991400	1.330000	12.460000

## ヒストグラム表示

```
> hist(dd12, 100)
```

```
> abline(v = 4.5, lty = 2)
```

## 4.5より右のものを数えてみると

```
> sum(dd12 >= 4.5)
```

```
[1] 27
```

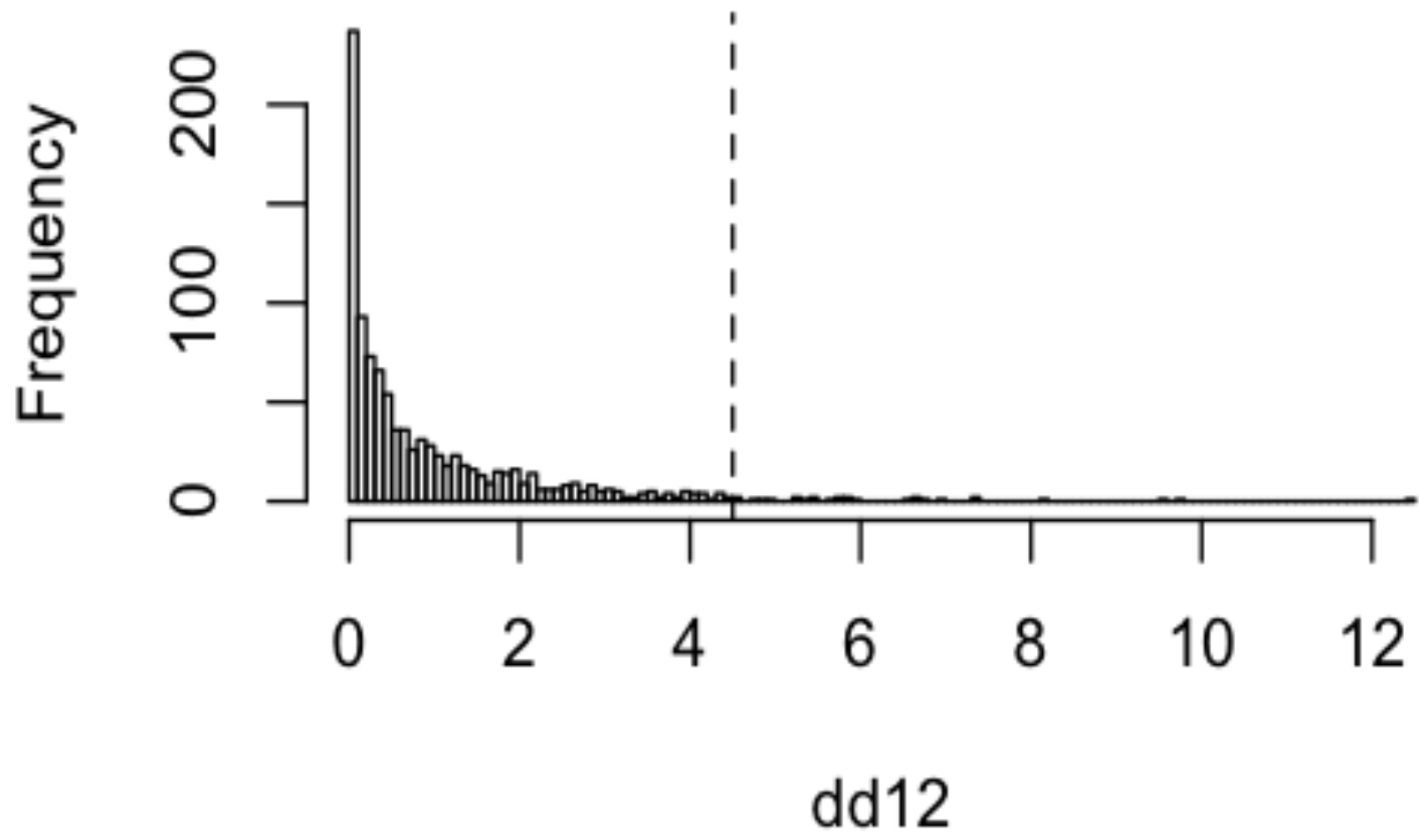
## P=0.05となる逸脱度の差を調べると

```
> quantile(dd12,0.95)
```

```
95%
```

```
3.566121
```

# ヒストグラム



# $\chi^2$ を使った近似計算法

PB法では、どんな統計モデルでも近似計算を使わないですむ。しかし、近似計算を使うと、もっとお手軽な場合がある。

fit1とfit2にモデルを格納し、anova()関数を使用。

```
> fit1 <- glm(y ~ 1, data = d, family = poisson)
> fit2 <- glm(y ~ x, data = d, family = poisson)
> anova(fit1, fit2, test = "Chisq")
```

Analysis of Deviance Table

Model 1: y ~ 1

Model 2: y ~ x

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1	99	89.507			
2	98	84.993	1	4.5139	0.03362 *

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1