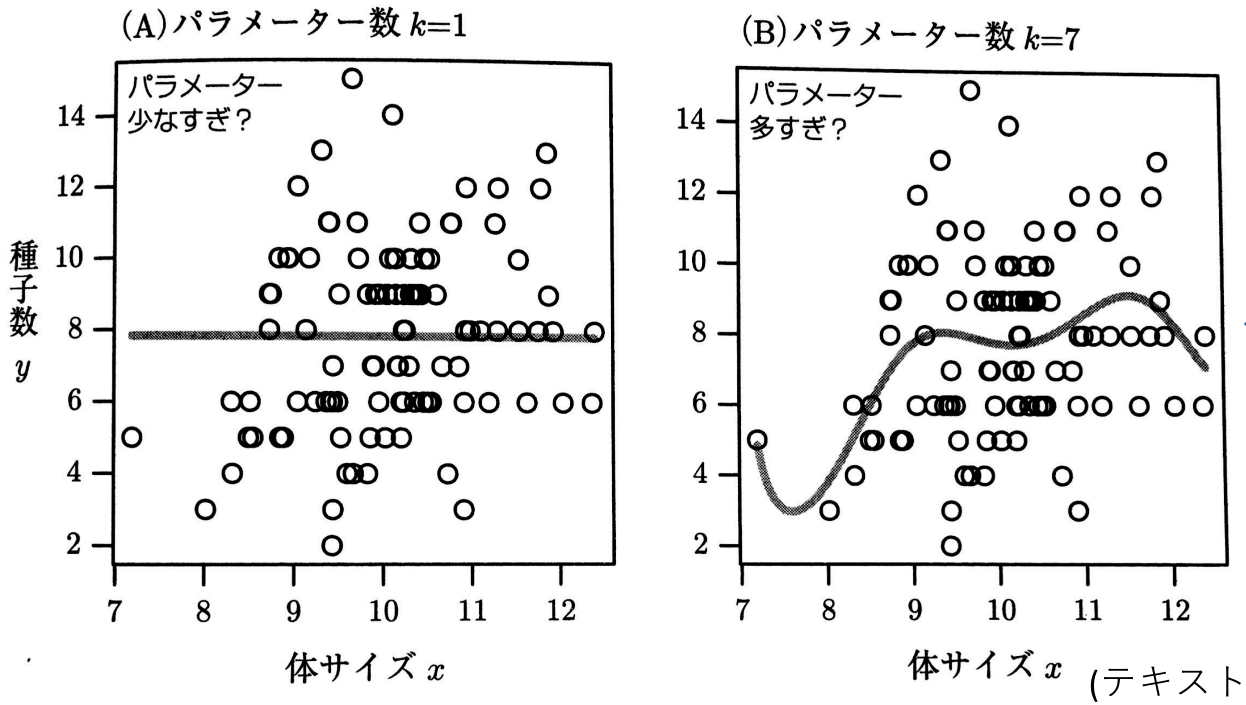
第4章:GLMのモデル選択(前半)

太田研究室 学部4年 和田

章の概要：「良い」モデルとは何か？どのような規準で選択すればいいか？

パラメーター数(k)を多くすればする分、**最大対数尤度**は大きくなる



(テキストより)

こちらの方がモデルとして優れている……とは限らない！

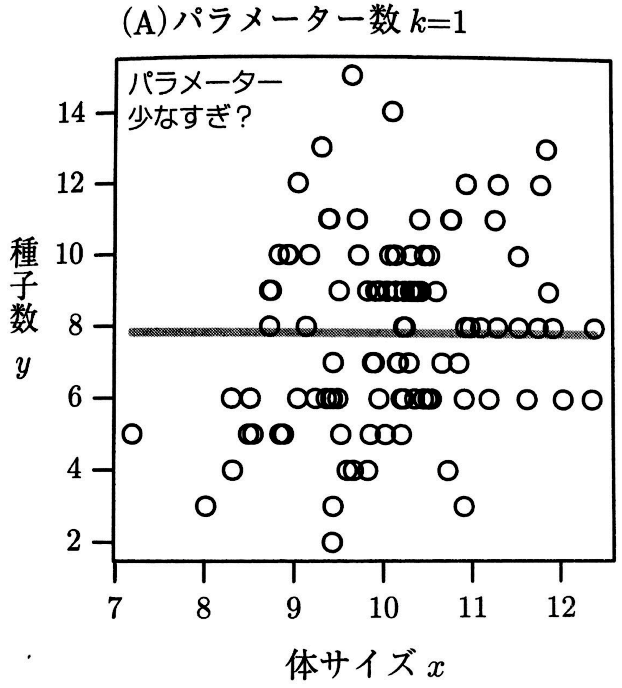
$$\log \lambda = \beta_1$$
$$k = 1$$
最大対数尤度：小

$$\log \lambda = \beta_1 + \beta_2 x + \ \dots + \beta_7 x^6$$
$$k = 6$$
最大対数尤度・大

章の概要：「良い」モデルとは何か？どのような基準で選択すればいいか？

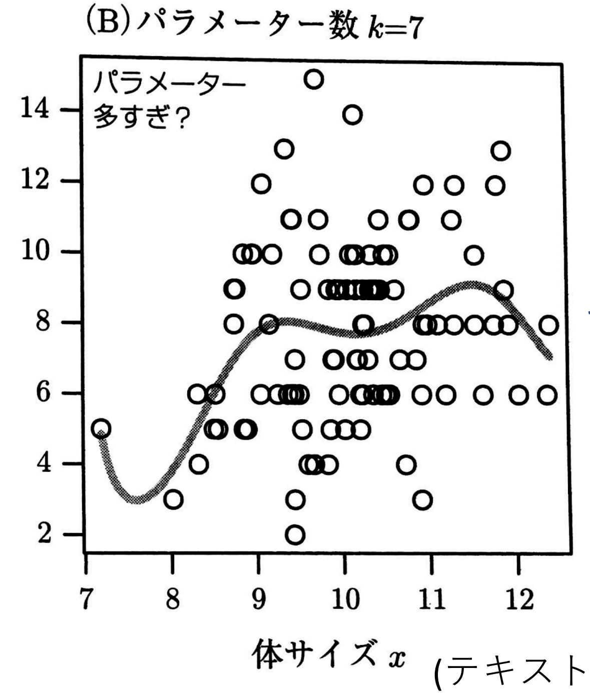
パラメーター数(k)を多くすればする分、**最大対数尤度**は大きくなる

観測データへのあてはまりの良さ



$$\log\lambda = \beta_1$$
$$k = 1$$

最大対数尤度：小



(テキストより)

$$\log\lambda = \beta_1 + \beta_2 x + \ldots + \beta_7 x^6$$
$$k = 6$$

最大対数尤度・大

こちらの方がモデルとして優れている

……とは限らない！

要確認

【理由】
・計算処理
・実際の現象との乖離

最大対数尤度以外の、新しいモデルの評価法・選択規準：
「当てはまりの悪さ」→**逸脱度**
「そのモデルは良い予測をするのか？」→ **AIC**

4.1 データはひとつ、モデルはたくさん

モデル候補の選定
「あてはまりの良さ」基準は万能ではない

逸脱度
「あてはまりの悪さ」という評価法

AIC
「予測の良さ」という選択規準

AICの有用性

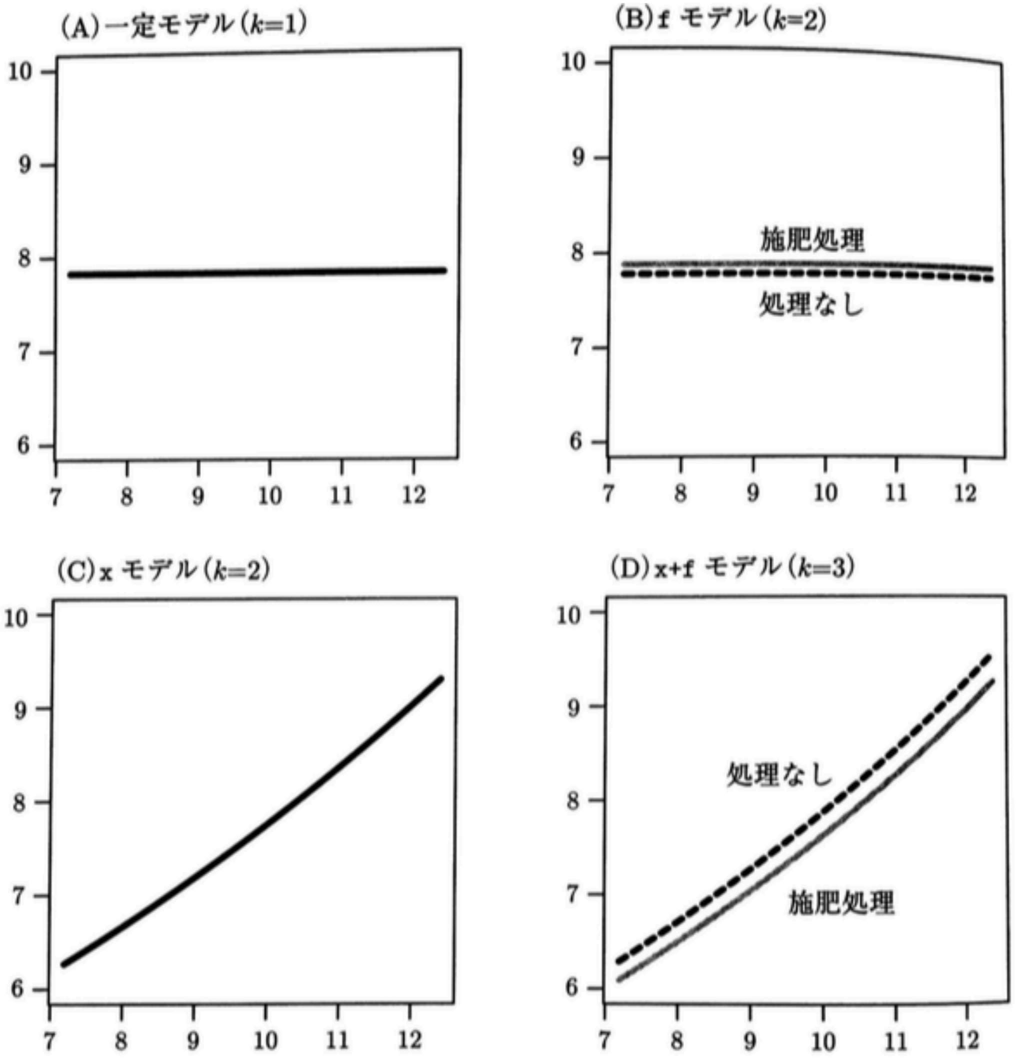
ネストの問題

なぜAICはモデル選択基準として有用なのか

一つのデータに対し、考慮する説明変数のパターン(=候補となるモデル)はたくさんある

体のサイズx_iも施肥の有無も、種子の量y_iに影響しない

施肥の有無f_iのみが種子の量y_iに影響する



体のサイズx_iのみが種子の量y_iに影響する

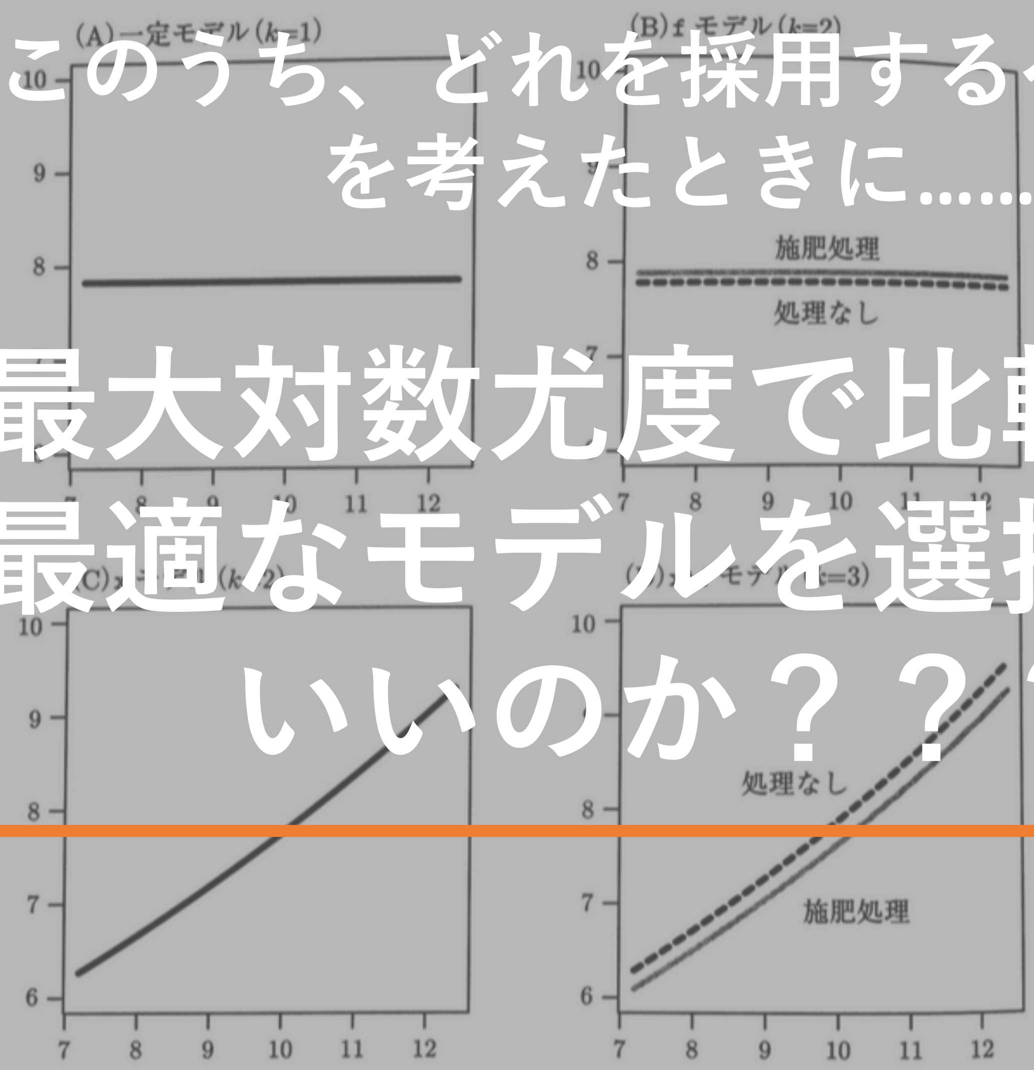
施肥の有無f_iと、体のサイズx_iの両方が、種子の量y_iに影響する

(テキストより)

一つのデータに対し、考慮する説明変数のパターン(=候補となるモデル)はたくさんある

体のサイズx_iも施肥の有無も、種子の量y_iに影響しない

体のサイズx_iのみが種子の量y_iに影響する



施肥の有無f_iのみが種子の量y_iに影響する

施肥の有無f_iと、体のサイズx_iの両方が、種子の量y_iに影響する

(テキストより)

このうち、どれを採用するべきか？
を考えたときに……

最大対数尤度で比較し、最適なモデルを選択していいのか？？？

4.2 統計モデルのあてはまりの悪さ：逸脱度

モデル候補の選定 「あてはまりの良さ」基準は万能ではない	**逸脱度** 「あてはまりの悪さ」という評価法	**AIC** 「予測の良さ」という選択規準	**AICの有用性** ネストの問題	 なぜAICはモデル選択基準として有用なのか

モデルのデータへのあてはまりの悪さ「逸脱度」は、最大対数尤度の変形

- 「**逸脱度**」(Deviance) D
- 統計モデルの、データへの「あてはまりの悪さ」の指標

$$D = -2\,logL^{*}$$

$logL(\{\beta_j\})$を$logL$、その最大対数尤度を$logL^{*}$と表記

glm()コマンドの出力結果に表示

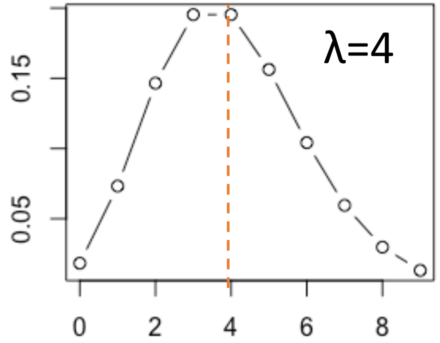
名前	In English	定義
逸脱度 (D)	Deviance	$-2\,logL^{*}$
最小の逸脱度	Minimum deviance	フルモデル(後述)をあてはめたときのD
残差逸脱度	$\boxed{\text{Residual deviance}}$	D – 最小のD
最大の逸脱度	Maximum deviance	Nullモデル(後述)をあてはめたときのD
Null 逸脱度	$\boxed{\text{Null deviance}}$	最大のD – 最小のD

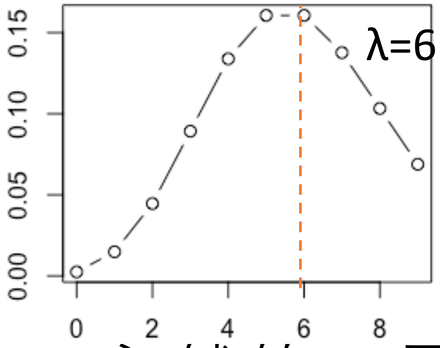
フルモデル、Nullモデルはそれぞれ、パラメーター数を最大、最小(1)にした場合のモデルである①

「**フルモデル**」(full model)　…　最もあてはまりが**いい**モデル

- 個々のデータに、一対一対応でパラメーター**λ**が定まっている
  - 100個のデータがあれば100個の**λ**を定めている

例：$y_i = \{4,4,4,6,\ldots\}$のとき、
$i \in \{1,2,3\}$の y_i は4なので、$\{\lambda_1, \lambda_2, \lambda_3\} = \{4,4,4\}$
$i = 4$のy_4は6なので、$\lambda_4 = 6$
…(以後同上)





- (同じ回帰で)他のどのモデルを使った時よりも、必然的に最大対数尤度は最大、逸脱度は最小になる

フルモデルを当てはめた時の
逸脱度＝最小のD(minimum deviance)

- 「現象を説明しうる理想のモデルを考えている」のではなく、「現在のデータ(のみ)にモデルを近づけている」ので、モデルとしての価値はない

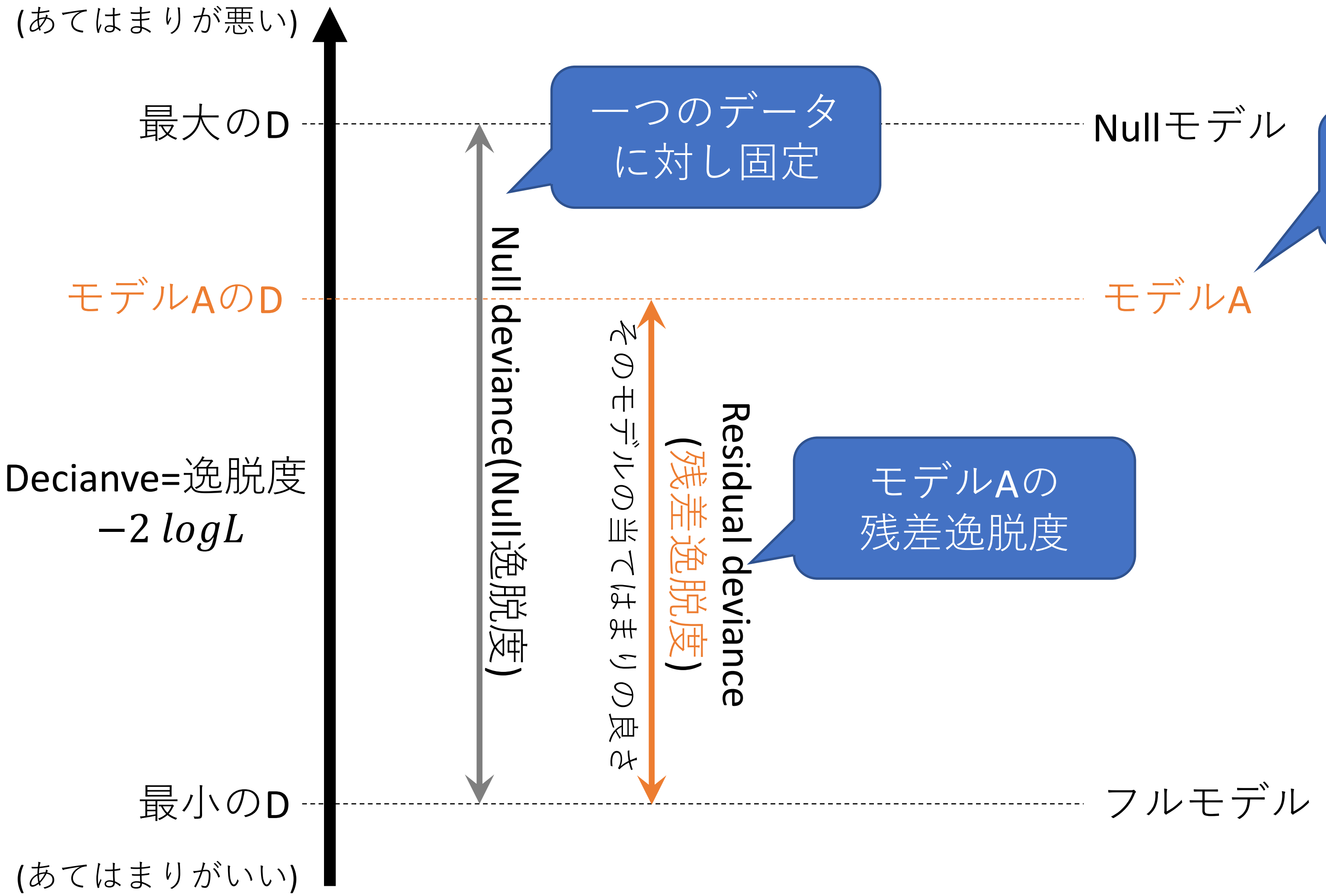
フルモデル、Nullモデルはそれぞれ、パラメーター数を最大、最小(1)にした場合のモデルである②

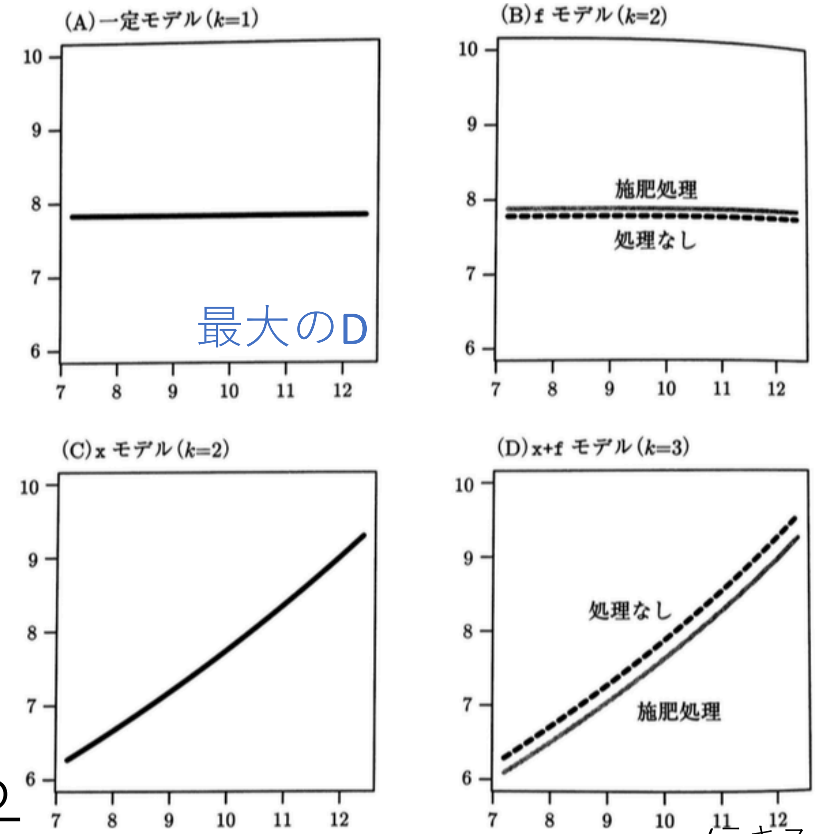
「**Null モデル**」(Null model)　…　最もあてはまりが**悪い**モデル

- パラメーター数が1
  - つまり、この文脈においては$\lambda = \exp(\beta_1)$
- パラメーターは、全ての説明変数から完全に独立である
- (同じ回帰で)他のどのモデルを使った時よりも、必然的に対数尤度は最小、逸脱度は最大になる

Null モデルを当てはめた場合の逸脱度
＝最大のD (Maximum deviance)

種々の逸脱度の関係性は以下



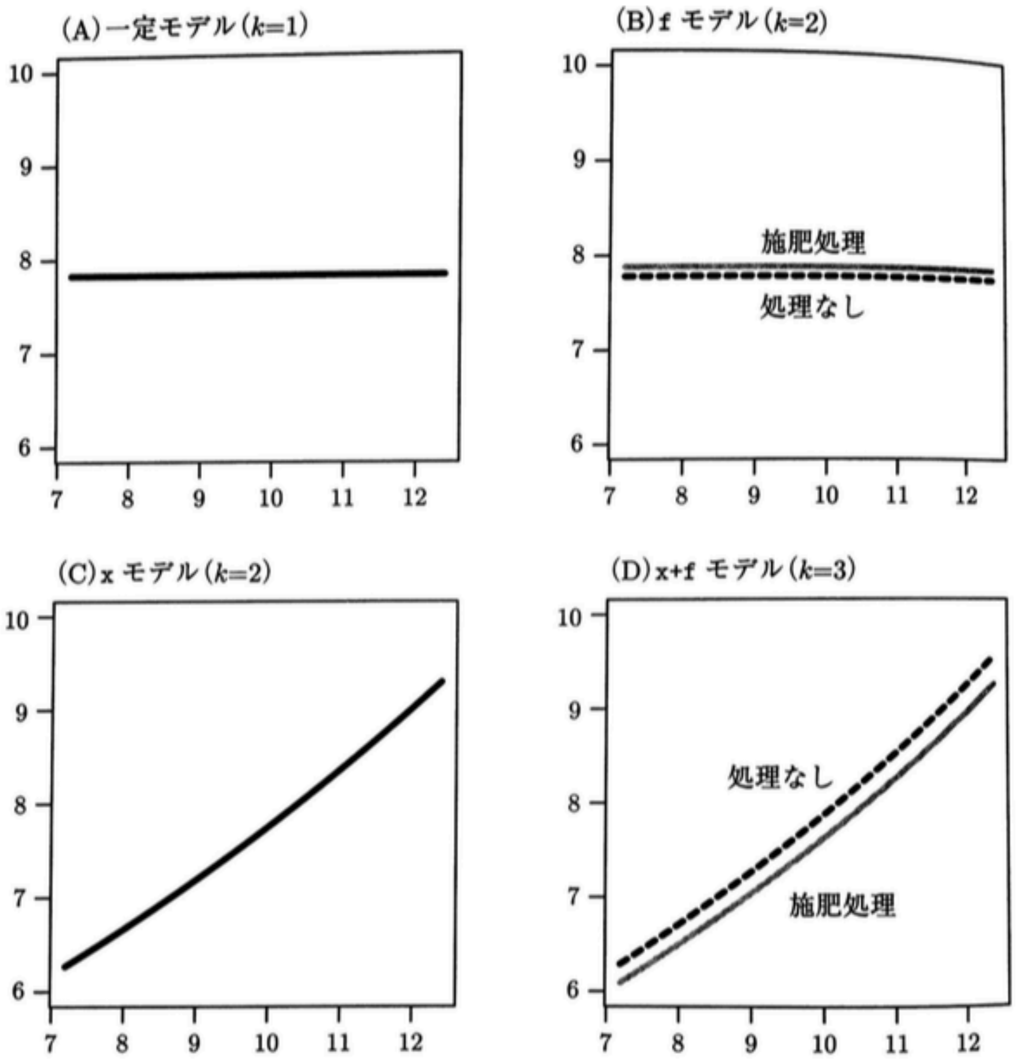


(テキストより)

<u>パラメーター数が増えるほど、(残差)逸脱度は減る</u>

R実践：残差逸脱度の算出・比較

R実践：残差逸脱度の算出・比較



(テキストより)

```
> fit.xf

Call:  glm(formula = y ~ x + f, family = poisson, data = d)

Coefficients:
(Intercept)            x           fT
    1.26311      0.08007     -0.03200

Degrees of Freedom: 99 Total (i.e. Null);  97 Residual
Null Deviance:       89.51
Residual Deviance: 84.81        AIC: 476.6
```

モデル	k	logL*	Deviance	Residual D
A:一定	1	-237.6	475.3	89.51
B: f	2	-237.6	475.3	89.48
C: x	2	-235.4	470.8	84.99
D: x+f	3	-235.3	470.6	84.81
フル	100	-192.9	385.8	0.0

4.3 モデル選択規準AIC

モデル候補の選定
「あてはまりの良さ」基準は万能ではない

逸脱度
「あてはまりの悪さ」という評価法

AIC
「予測の良さ」という選択規準

AICの有用性

ネストの問題

なぜAICはモデル選択基準として有用なのか

AICの比較により「予測の良さ」を重視したモデル選択を行うことができる

AIC (Akaike's information criiterion)

- 「モデル選択規準」(model selection criterion)の一つ
- 予測の良さを重視する　　(~~当てはまりの良さ~~)
- 小さい方が「良い」モデル

$$AIC = -2\{(\text{最大対数尤度}) - (\text{パラメーター数})\}$$
$$= -2(logL^* - k)$$
$$= \boxed{-2logL^*} + 2k$$

要確認

「基準」と「規準」の違いとは？

「**規準**」…何を測定するか。何の指標を用いるか。
辞書「何かを行う際に手本・標準とすべきもの」
Eg「道徳の規準」「社会生活の規準」

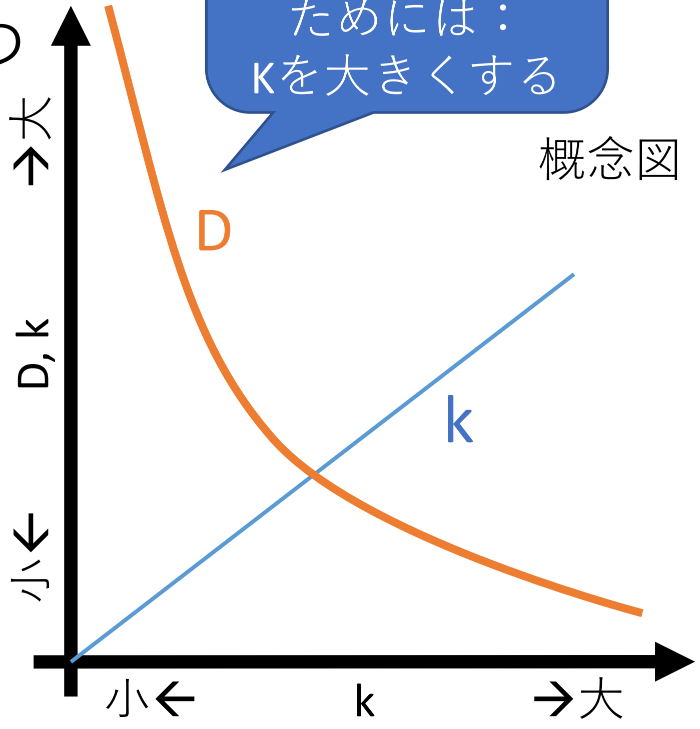
「**基準**」…どこまで達成できたか。測定された値に基づく評価。
辞書「物事を判断するためのよりどころ。標準と見なす数値など」
Eg「選考基準」「前年を基準に予算を決める」

AICの比較により「予測の良さ」を重視したモデル選択を行うことができる

AIC (Akaike's information criiterion)

- 「モデル選択規準」(model selection criterion)の一つ
- 予測の良さを重視する　　(~~当てはまりの良さ~~)
- 小さい方が「良い」モデル

ジレンマ
Dを小さくするためには：
Kを大きくする



$$AIC = -2\{(\text{最大対数尤度}) - (\text{パラメーター数})\}$$
$$= -2(logL^* - k)$$
$$\boldsymbol{= D + 2k}$$

これを小さくしたい

右辺の二項、両方を小さくしたい

逸脱度とパラメーター数の両方が小さい→「良い」モデル

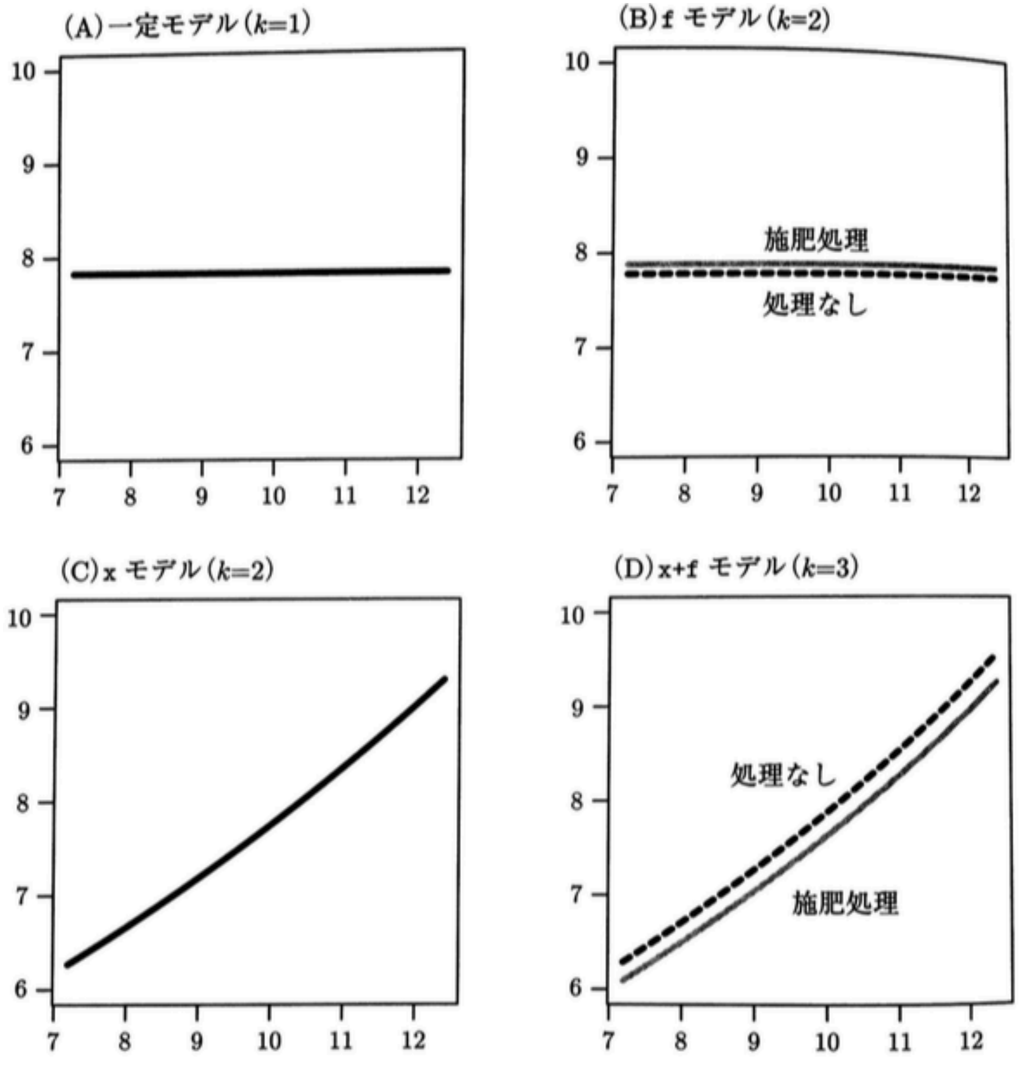
疑問

逸脱度の概念は必要か？最大対数尤度じゃダメなのか？
残差逸脱度の概念が必要なのは分かるのだが…

(AICが選択規準として有用である理由は以後)

R実践：AICの算出・比較

R実践：AICの算出・比較



(テキストより)

```
> fit.xf

Call:  glm(formula = y ~ x + f, family = poisson, data = d)

Coefficients:
(Intercept)            x           fT
    1.26311      0.08007     -0.03200

Degrees of Freedom: 99 Total (i.e. Null);  97 Residual
Null Deviance:      89.51
Residual Deviance: 84.81        AIC: 476.6
```

モデル	k	logL*	Deviance	Residual D	AIC
A:一定	1	-237.6	475.3	89.5	477.3
B: f	2	-237.6	475.3	89.5	479.3
C: x	2	-235.4	470.8	85.0	474.8
D: x+f	3	-235.3	470.6	84.8	476.6
フル	100	-192.9	385.8	0.0	585.8

残差逸脱度がDより大きいのに、AICはCが最小

ここまでのまとめ

- **最大対数尤度**は「あてはまりの良さ」を表現している
- 最大対数尤度は、モデルの良さの評価規準としてふさわしいのか？？
  - 「一つのデータセット」の説明に終始→実際の現象と乖離する可能性が高い
- **逸脱度**＝「あてはまりの悪さ」という値でモデルを評価可能
- **残差逸脱度**は、他のモデルの逸脱度と比較した相対的な規準である
  - **フルモデル**採用時の逸脱度=最小のDとの比較
- **モデル選択規準 AIC**の値が小さいほど「良い」モデルと言える
- AICは「モデルの予測の良さ」を判断規準とする
- パラメーター数kと逸脱度Dの双方が程よく小さい時：AICが最小
- 「良い予測をする」モデルは、kとDが小さい

4.4 AICを説明するためのまた別の問題

モデル候補の選定
「あてはまりの良さ」基準は万能ではない

逸脱度
「あてはまりの悪さ」という評価法

AIC
「予測の良さ」という選択規準

AICの有用性
ネストの問題
なぜAICはモデル選択基準として有用なのか

モデルAがモデルBの部分集合であるとき、モデルA,Bは「**ネストしている**」

<u>「**ネストしている/した**」</u>(nested)

- 一方のモデルが他方に含まれている状態

$$\text{モデルA:} \quad \log\lambda = \boxed{\beta_1 + \beta_2 x_1}$$
$$\text{モデルB:} \quad \log\lambda = \boxed{\beta_1 + \beta_2 x_1} + \beta_3 x_2$$

- 上記の場合、モデルAはモデルBの部分集合
  ＝「モデルAとBはネストしている」
- どんなモデルも、フルモデルの部分集合
  ＝フルモデルは、必然的にどんなモデルともネストしている
- ネストしたモデル間の検定：Wald検定、尤度比検定など

4.5 なぜAICでモデル選択して良いのか？

- **モデル候補の選定**
  「あてはまりの良さ」基準は万能ではない
- **逸脱度**
  「あてはまりの悪さ」という評価法
- **AIC**
  「予測の良さ」という選択規準
- **AICの有用性**
  - ネストの問題
  - なぜAICはモデル選択基準として有用なのか