

一般化線形混合モデル (Generalized Linear Mixed Model)

個体差や場所差の影響をGLMにくみこんだ統計モデル

個体差をあらわすパラメーター r_i が追加される

$$\text{logit}(q_i) = \beta_1 + \beta_2 x_i + r_i$$

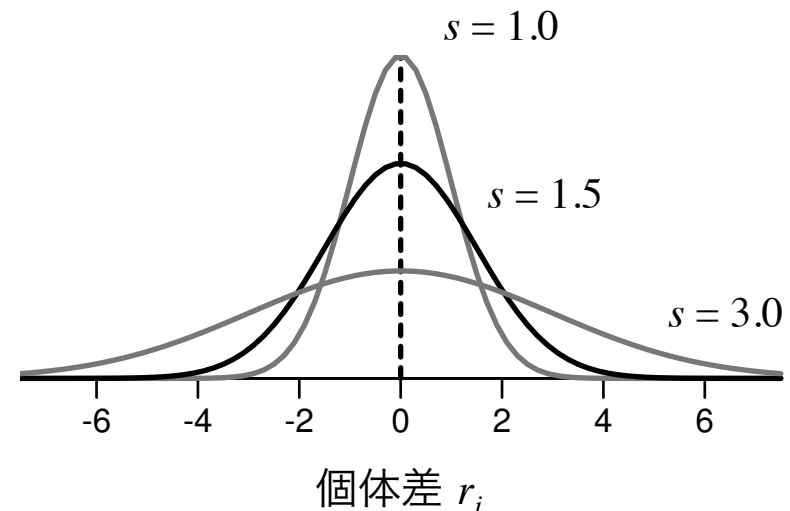
推定値自体の大きさには影響を与えずに、推定のばらつきだけに影響する

r_i は $-\infty$ から $+\infty$ までの範囲をとる連続値で平均ゼロ、標準偏差 s の正規分布にしたがう

r_i の確率密度関数 (出現しやすさ) は：

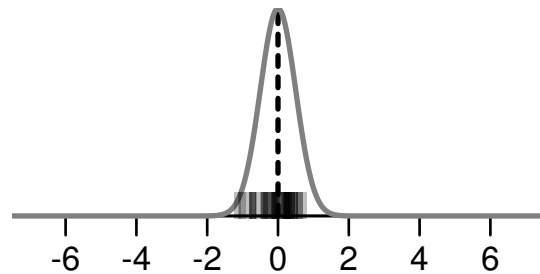
$$p(r_i | s) = \frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{r_i^2}{2s^2}\right)$$

標準偏差 s は「集団内の r_i のばらつき」をあらわす

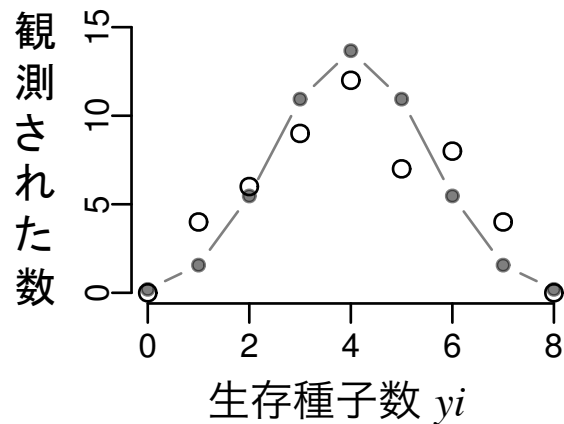


個体差 r_i のばらつき s が大きいと過分散になる

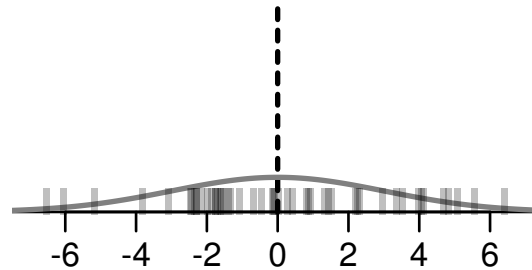
個体差のばらつきが小さい場合
($s = 0.5$)



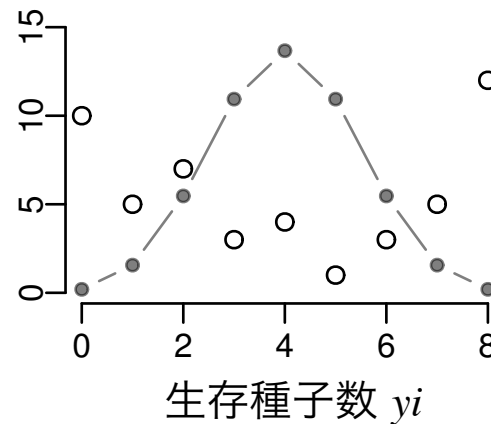
個体差 r_i



個体差のばらつきが大きい場合
($s = 3.0$)



個体差 r_i



- $s = 0$ (全個体均質) と仮定したときのモデルの予測
- 観測された生存種子数ごとの頻度

一般化線形混合モデル (Generalized Linear Mixed Model)

固定効果(fixed effect)とランダム効果(random effect)の混合(Mixed)モデル

$$\text{logit}(q_i) = \underbrace{\beta_1 + \beta_2 x_i}_{\text{固定効果}} + \underbrace{r_i}_{\text{ランダム効果}}$$

固定効果

→最尤推定する

ランダム効果

→最尤推定しない (できない)

最尤推定のやり方

$$L_i = \int_{-\infty}^{\infty} p(y_i | \beta_1, \beta_2, r_i) p(r_i | s) dr_i$$

- (1) いろいろな r_i における尤度を評価して期待値を算出し、
- (2) 各 r_i の出現しやすさ「 $p(r_i | s)$ 」で重み付け、
- (3) それらを足し合わせる (分布の混ぜ合わせ)
→過分散な確率分布を作り出せる

Rを使ってGLMMのパラメーターを推定しよう

カウントデータのGLMMの最尤推定なので、`glmmML` パッケージを使う

`glmmML` パッケージはRの標準パッケージではないので、CRANサイトからダウンロードする

ダウンロードできたら、`library(glmmML)`と指示してパッケージを読み込む

GLMMのコードを走らせる：

```
glmmML ( cbind (y, N - y) ~ x, data = d, family =binomial, cluster = id)
```

	coef	se(coef)	z	Pr(> z)
(Intercept)	-4.190	0.8777	-4.774	1.81e-06
x	1.005	0.2075	4.843	1.28e-06

Scale parameter in mixing distribution: 2.408 gaussian

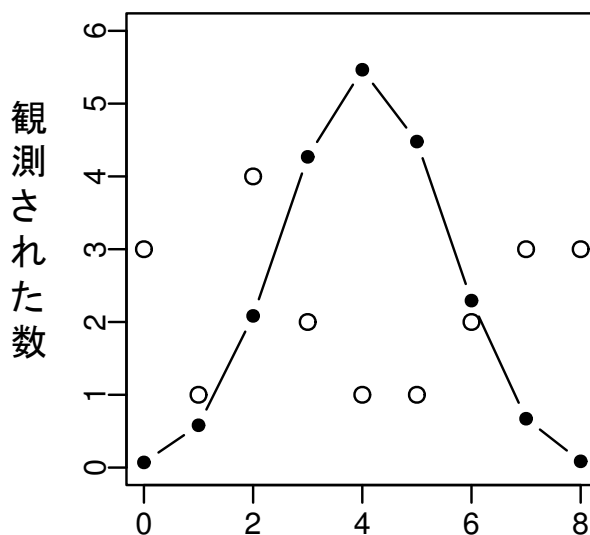
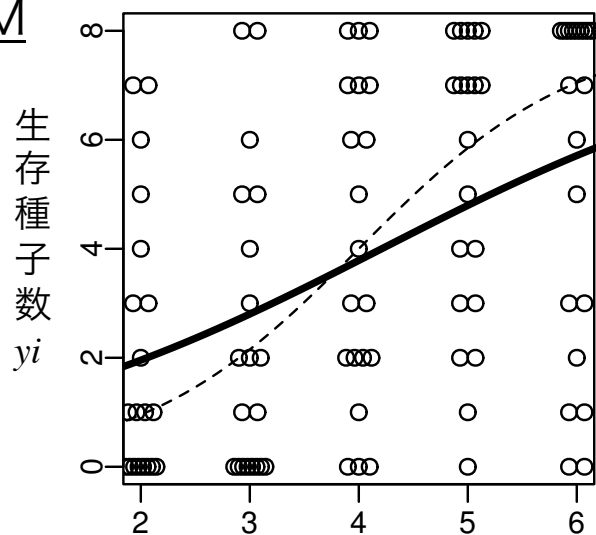
Std. Error: 0.2202

Residual deviance: 269.4 on 97 degrees of freedom AIC: 275.4

こんな感じの結果になる…はず

GLMMを使うことで推定が改善

GLM



----- 「真の」生存確率
 ——— 推定の生存確率

○: 観測された種子数分布
 ●: 推定された種子数分布

「真の」値: $\beta_1 = -4$
 $\beta_2 = 1$

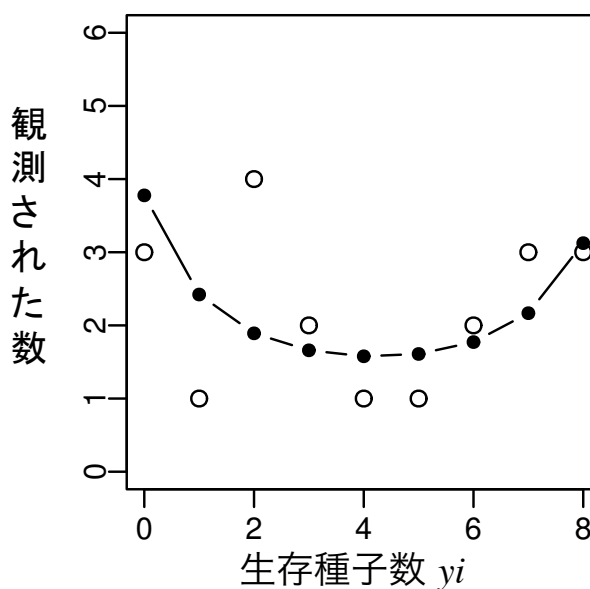
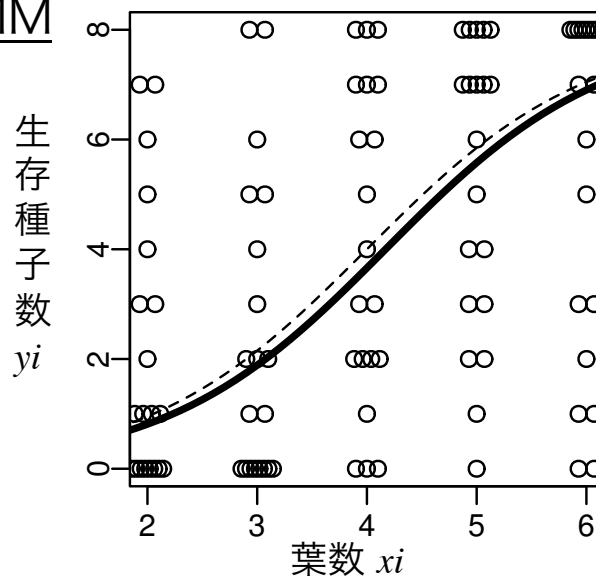
GLMの推定値:

$\beta_1 = -2.15$
 $\beta_2 = 0.51$

GLMMの推定値:

$\beta_1 = -4.19$
 $\beta_2 = 1.005$

GLMM



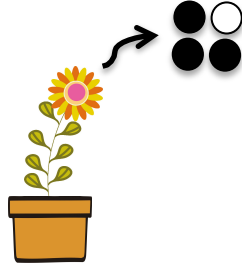
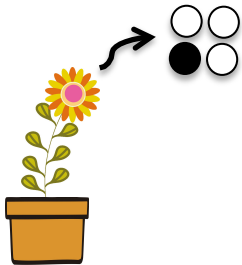
反復・疑似反復 (pseudo replication) と統計モデルの関係



植木鉢が反復

個体差も植木鉢差も推定できない

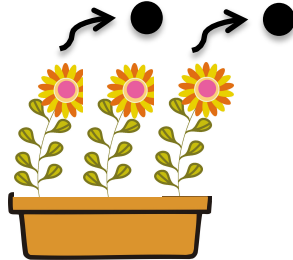
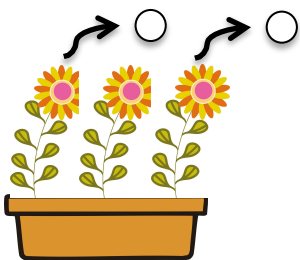
$$\text{logit}(q_i) = \beta_1 + \beta_2 x_i$$



個体は疑似反復、植木鉢は反復

個体差は推定できる、植木鉢差は推定できない

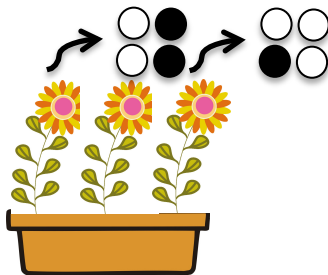
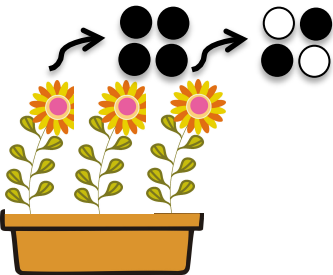
$$\text{logit}(q_i) = \beta_1 + \beta_2 x_i + r_i$$



個体は反復、植木鉢は疑似反復

個体差は推定できない、植木鉢差は推定できる

$$\text{logit}(q_i) = \beta_1 + \beta_2 x_i + r_j$$



個体も植木鉢も疑似反復

個体差は推定できない、植木鉢差は推定できる

$$\text{logit}(q_i) = \beta_1 + \beta_2 x_i + r_i + r_j$$

→階層ベイズモデル

いろいろな分布のGLMM

応答変数のばらつき

二項分布
ポアソン分布 } → glmmML() をつかう

正規分布
ガンマ分布 } → lme4パッケージのglmer() などをつかう

※ランダム効果の指定のしかたが違うので注意！

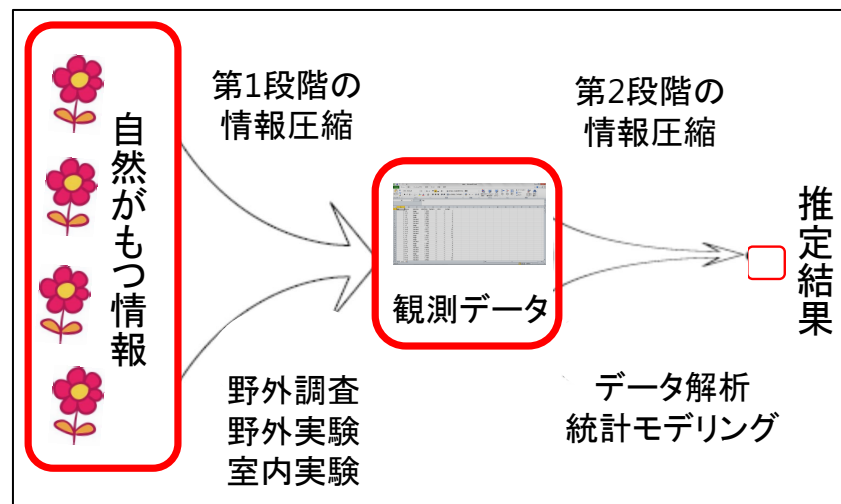
例：

```
result<-glmer(attacked_order~prey_group+(1|position),  
data=data)
```

まとめ

- 観測されていない個体差があるために、現実のデータの分布はGLMで期待されるより過分散なものになる
- GLMMは、線形予測子に個体差のばらつきをあらわすパラメーター r_i を追加することで過分散に対応できる
- 一個体から複数のデータをとったり、一つの場所に多数の調査対象がいるような状況は疑似反復とよばれ、個体差・場所差などをくみこんだGLMMを使う必要がある
- 確率分布の種類によらず、個体差・場所差などに影響されるデータの部分集合があれば、これらをランダム効果としてくみこんだ統計モデルを使わなければならない

さらなる発展



階層ベイズモデル

MCMCによる事後分布の推定

個体差&場所差
空間相関などより複雑な
要因を組み込んだ
統計モデルをあつかいたい

一般化線形混合モデル

個体差・場所差
といったランダム
効果をあつかいたい

一般化線形モデル

最尤推定法

最小二乗法

線形モデル

正規分布以外の確率
分布をあつかいたい