データ解析のための統計モデリング (2008 年 10-11 月)

全 5 (+2) 回中の第 7 回 (2008–12–03)

階層ベイズモデル

久保拓弥 kubo@ees.hokudai.ac.jp

http://hosho.ees.hokudai.ac.jp/~kubo/ce/EesLecture2008.html

この講義のーとが「データ解析のための統計モデリング入門」として出版されました!

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

まちがいを修正し，詳しい解説・新しい内容をたくさん追加したものです

今日のもくじ

1. ベイズ統計って …… ようするに「便利な道具」 2
2. 今日はすごく単純化した例題で 4
3. 最尤推定法の復習 5
4. GLM では説明できない個体差由来の過分散 6
5. 統計モデルに個体差パラメーターをいれてみる 8
6. 階層ベイズモデルで表現する個体差 8
7. 経験ベイズ法による最尤推定，すなわち GLMM 11
8. Markov chain Monte Carlo 法とは何か? 13
9. 階層ベイズモデルの MCMC 計算 15
10. 事後分布として表現される推定結果 17

……「高次元母数の推定は頻度主義よりベイズがよい …… この点は認めざるをえない」「頻度論にもとづく仮説検定はですね，強い尤度原理を認めてないんですよ!」「MCMC は Bayesian ではない，単なる積分だ」「MCMC が Bayesian を普及させた」「あのヒトは真正 Bayesian ではない …… Bayesian をわかってないんだから」「頻度論的な統計学を正しく使いこなせるのは Fisher のような天才だけ」「Bayesian は理解できてないバカが使っても間違いがない …… それが Bayesian の良いところです」「Fisher 流の有意差検定と Neyman 流の仮説検定はまったく別モノ」「Bayes と Fisher は意外と似ている …… 違いは事前分布の有無だけ」「Neyman は違う，確率の考えかたがぜんぜん異なる」「無情報事前分布? そりゃー臆病だよ，あんたは臆病だ!」……

2005 年 9 月 13 日，統計関連学会連合広島大会，
「科学的な推論の形式としての Bayes 統計」セッションにおける
いささか興奮ぎみの統計学者たちの議論を傍聴していた久保による偏った記録
ぎょーむ日誌 http://hosho.ees.hokudai.ac.jp/~kubo/log/2005/0911.html#03 より

「生態学の統計モデリング」第 7 回目，よーやく最終回です．[1] さてさて，今日のハナシですが……前回説明した一般化線形混合モデル (GLMM) データ解析に使えるようになると，今度は観測データのあちこちに random effects になりうる要因があることに気づきます．たとえば野外観測データの場合，「場所差」があってその中に「個体差」がある，といった状況です．

1. 2008 年度の統計学授業，前回・今回は熱心なる皆さんのご希望により，「補講」として実現してしまいました……

このように統計モデルの中で複数の random effects をあつかう必要が生じると，最尤推定法によるパラメーター推定がだんだん**苦しく**なります．そこで，GLMM を**階層ベイズモデル**の一種だとみなして，[2] **Markov Chain Monte Carlo (MCMC) 法**を使ってパラメーターの事後分布をもとめる方法を紹介します．

2. つまり前回は「GLM を拡張すると GLMM になる」という説明にしましたけれど，今日は「階層ベイズモデルを簡単化すると GLMM になりますよ」という方向で説明します．

1. ベイズ統計って …… ようするに「便利な道具」

ここまで，この統計学授業であつかってきたのは，**とくに指摘しませんでしたが**，非ベイズすなわち**頻度主義** (frequentism) な統計学の道具だてと考えられています．で，歴史的には非ベイズとベイズな人たちはキビしく対立してきた経緯があり— いまでも原理的にはあいいれない対立なのですが，とりあえずわれわれにはあまり**関係ありません**．というのも私たちは科学哲学の研究者ではなく，自然科学の研究者なので，研究に使う道具だてとしては次の条件を満たしていれば，とりあえずは「使える」からです．

1. 数学的にきちんとしている，つまり形式上の矛盾があからさまではない
2. できるだけ「客観的っぽい」，つまり恣意的な点が露骨ではない道具 [3]

3. もちろん字義どーりの意味で「客観的」(ぜんぜん恣意的でない) 自然科学の研究はありえません．

ベイズ統計モデリングはいまや上の二点を満たしているので，われわれは道具として使うことができるわけです．

まずは「ベイズ統計学とは何か?」そのおハナシだけ先に簡単にすませてしまいましょう．「ベイズ統計学って何?」というおおざっぱな質問に対するおおざっぱな回答は以下のとーりです:

- **主観確率** をあつかう，[4] つまり「明日の降水確率は 20%」の解釈は「雨のふる・ふらないは 2:8 ぐらい，といった割合でおこりそうな事象だと考えている」となる
  - これに対して頻度主義な統計学では，**客観確率** をあつかう，つまり「明日の降水確率は 20%」の解釈は「明日という日が無限個存在していたとすると，そのうち 20% では雨が降っていると観測される」となる
- 確率分布の**パラメーターもまた確率分布**として表現される

4. 主観確率・客観確率のどちらも公理主義的確率論が要請する条件を満たしています— つまり確率論のさまざまな道具だてを援用できます．つまるところ主観確率・客観確率という区別は,「確率って何なの?」という確率の解釈に関する相違なのです．

- この「パラメーターの確率分布」には**事前分布** (prior) と**事後分布** (posterior) がある
- 事前分布は「パラメーターの値はこのへんにありそうだな」をあらわす確率分布，解析者が決める; ただし実際のデータ解析では (後述するように) 次のふたつの事前分布をよくつかう
  - 無情報事前分布: 「パラメーターの値はどこにあってもいいんですよ!」ということを表現する事前分布，つまり「べたーっとひらべったい事前分布」— fixed effects 的な効果を表現するパラメーターの事前分布などとして使う
  - 階層モデル的な事前分布: 事前分布のカタチを観測データと超パラメーター (hyper parameter) で決める，超パラメーターは無情報事前分布をもつ— random effects 的な効果を表現するパラメーターの事前分布などとして使う
- 事後分布は「ベイズ統計モデル」と観測データから**推定**される— つまり観測データから事後分布を推定することがベイズ推定 (Bayesian inference) の目的である
- 上でいう「ベイズ統計モデル」は次のふたつの部品からなる
  - 尤度(ゆうど): つまり統計モデルが観測データにどれぐらいあてはまっているか，を確率分布とそのパラメーターであらわす確率 (確率密度) — これまでこの授業でとりあつかってきた尤度と同じ
  - 事前分布 (と超事前分布): 尤度にふくまれるパラメーターをあらわす確率分布 (確率密度関数)

ついでに，頻度主義な統計学的ツールとの比較のテイブルでもこしらえてみましょう.

	頻度主義な道具	**ベイズな道具**
確率とは?	客観確率	主観確率
パラメーターとは?	ある**特定の値**，世界のどこかに「真の値」があるはず	**確率分布**で表現できることにする
事前分布は?	ない (強いていえば無限の幅をもつ一様分布)	推定すべきパラメーターにあわせて様々な事前分布を設定する
観測データ使って何を推定する?	「真の値」に近いはずのパラメーター**最尤推定値**，つまりひとつの値 (点推定値)	パラメーターの**事後分布**，つまり確率分布
よく使う推定計算方法は?	数値的な**最尤推定法**，など	**MCMC 法**など; 階層ベイズモデルの最尤推定は「経験ベイズ法」と呼ばれる
推定されたパラメーターの信頼区間 α の意味は?	じつは難解	簡単，そのパラメーターが確率 α でとりうる範囲

えー，つまり今回のハナシで明らかにしたい点としては……非ベイズ vs ベイズな歴史的対立なんかがあったにもかかわらず，いまや自然科学の道具

としてベイズ統計モデリングは「非ベイズなわくぐみではうまく表現できなかった現象 (観測データの中にみられるさまざまなパターン) をうまく表現できるように**強化された統計モデリング技法**」にすぎない，といったあたりです.

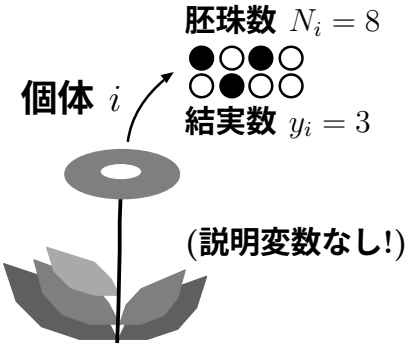
推定したい事後分布の「もと」になる事前分布の決めかたが「客観的」でなかったので**アヤしい**ものだと考えられていました. [5] しかしながら (最尤法や GLM の普及の歴史と同じく) 計算機ハードウェア・ソフトウェアの発達によって,「客観的っぽい」**階層ベイズモデル** (hierarchical Bayes model; HBM) による推定計算が可能になり，20 世紀末から急速に応用分野を拡大しているところです.

5. 以前のベイズ統計モデルな人たちは,「ユーザーが自由に事前分布を決めてよい」という点ばかりを強調されていて，ですね……われわれ科学研究者は「他人の目」を気にしながらこそこそ生きているので「自分で好き勝手に決めればぁ?」と言われると不安で落ちつかないキモチになっていたわけです.

2. 今日はすごく単純化した例題で

今日は今までの講義全体を復習するようなかんじでハナシをすすめていって，最終的に階層ベイズモデルにつながるようにしましょう. 前回の GLMM 例題の架空植物と同じような結実確率を推定する例題ですけど，今回はさらに簡単化されています.

ここではこの架空植物の個体ごとの (種子数ではなく) **結実確率**がどのように決まるか (統計モデルでどう表現するのがよいのか) をあつかいたいとします. いつもと同じく，個体は i という記号であらわされ ($i = 1, 2, 3, \cdots, 100$, つまり 100 個体います)，その胚珠 (はいしゅ) 数は 8 個 (全個体共通)，結実した種子数は y_i とします. 今回はサイズだの肥料だのサイズだの，といった説明変数になりそうな要因はありません!



いつものごとく，とりあえずデータを `read.csv()` してみましょう. `id` 列は個体番号 ($i = 1, 2, \cdots, 100$) です.

```
> d <- read.csv("data7a.csv")
> head(d)
  id y
1  1 0
2  2 2
3  3 7
4  4 8
5  5 1
6  6 7
> summary(d)
       id               y
 Min.   :  1.00   Min.   :0.00
 1st Qu.: 25.75   1st Qu.:1.00
 Median : 50.50   Median :4.00
 Mean   : 50.50   Mean   :4.03
 3rd Qu.: 75.25   3rd Qu.:7.00
 Max.   :100.00   Max.   :8.00
```

3. 最尤推定法の復習

今回は説明変数がない単純な例題なので，GLM をもちだすまでもなく最尤推定法で統計モデリングできそうです……ただし「個体差はない」と仮定する必要がありますけど．

ついでに，二項分布のパラメーター推定の最尤（さいゆう）推定を復習しましょう．**すべての個体で結実確率 q が共通している**という仮定のもとでは，個体 i の 8 胚珠の中で結実した胚珠数が y_i 個となる確率は二項分布

$$f(y_i \mid q) = \binom{8}{y_i} q^{y_i}(1-q)^{8-y_i},$$

で表現できます．植物 100 個体の観察値 $\{y_i\} = \{y_1, y_2, \cdots, y_{100}\}$ が観察される確率は上の $f(y_i \mid q)$ を 100 個体ぶん掛けあわせたものになります．このときに，逆に観察データ $\{y_i\}$ が与えられたもので，パラメータ q は値が自由にとりうると考えると，この 100 個体ぶんの確率はパラメータ q の関数となります．これは尤度とよばれ，形式的には

$$L(q \mid \{y_i\}) = \prod_{i=1}^{100} f(y_i \mid q)$$

と定義されます．この尤度 $L(q \mid \{y_i\})$ を最大化するパラメータの推定量 $\hat{q}$ を計算してみましょう．尤度を対数尤度になおして，

$$\log L(q \mid \{y_i\}) = \sum_{i=1}^{100} \log \binom{8}{y_i} + \sum_{i=1}^{100} \{y_i \log(q) + (8-y_i)\log(1-q)\}$$

この q に関する微分がゼロになる q を計算すると，$\hat{q} = \sum_{i=1}^{100} y_i/800$ つまり「結実した全胚珠個数」を全胚珠個数 (800) で割った割り算値が最尤推定量になっています．[6] 最尤推定値[7] は

```
> (q <- sum(d$y) / 800) # 割り算で得た結実確率の最尤推定値
[1] 0.50375
```

結実確率の最尤推定値は $\hat{q} = 0.504$ ぐらいになったので,

```
> 8 * q # 一個体内でいくつぐらい結実するかという期待値
[1] 4.03
> mean(d$y) # とうぜんながら実際の平均値もこうなる
[1] 4.03
```

と平均値に関しては二項分布を使った統計モデルで説明できてるように見えます.

ついでながら，R の glm() を使ったとしても

```
> glm(cbind(y, 8 - y) ~ 1, family = binomial, data = d)
...(略)...
Coefficients:
(Intercept)
      0.015
...(略)...
> 1 / (1 + exp(-0.015)) # beta の推定値から q を計算
[1] 0.5037499
```

とまったく同じ推定値が得られます.

6. あとで自分でやってみてください.

7. estimator と estimate のちがいは覚えていますか?

4. GLM では説明できない個体差由来の過分散

しかしながら,「より良い統計モデリング」[8] では，**データのばらつき**つまり分散にも注意をはらいます．$N = 8$ かつ $q = 0.504$ の二項分布で期待される分散と実際の分散を比較してみると

```
> 8 * q * (1 - q) # 二項分布で期待される分散
[1] 1.999888
> var(d$y)
[1] 9.928384
```

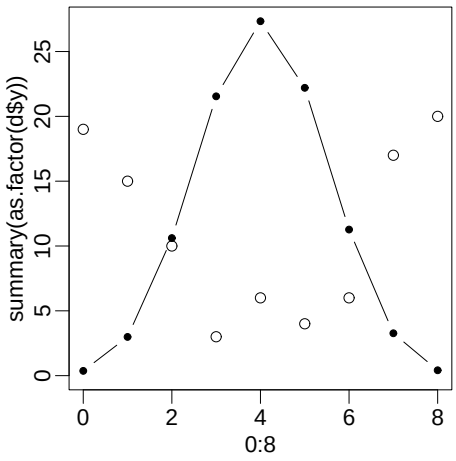
このようにあきらかに過分散が生じている，とわかりました.「結実数 y 個の個体数は? (合計 100 個体)」を調べてみましょう:

8. つまりよりよいデータ解析

```
>table(d$y) # 結実数 y 個の個体数は?（合計 100 個体）
 0  1  2  3  4  5  6  7  8
19 15 10  3  6  4  6 17 20
> round(dbinom(0:8, 8, q) * 100, 1) # 二項分布の予測
[1]  0.4  3.0 10.6 21.5 27.3 22.2 11.3  3.3  0.4
```

結実確率 $\hat{q} = 0.504$ の二項分布モデルの予測と観察データを比較すると，ぜんぜんくいちがっています．[9] [10] 図で確認してみればすぐにわかるのですが

```
> plot(0:8, table(d$y))
> lines(0:8, dbinom(0:8, 8, q) * 100, type = "b", pch = 20)
```



これを見ると，結実する確率 q は平均値計算で推定してしまえばよい，とする二項分布モデルでは現象をうまく説明できていない，と見当がつきます．[11] おそらく,「どの個体でも胚珠が結実する確率 q は同じ (この例だと 0.5 ぐらい?)」という二項分布の仮定があまり正しくないのでしょう．このように，個体 i の結実種子数 y_i のばらつきが二項分布モデルの予測から逸脱してしまう現象を過分散 (overdispersion) とよびます．

観察データの過分散パターンを説明するためには，二項分布モデルを拡張しなければなりません．たとえば，**結実する確率 q は植物個体によって異なる**らしいと考えてみるのは自然なことです．個体ごとに結実確率が集団平均からずれていることを，仮に個性もしくは個体差とよぶことにします．

なぜ個体に差が生じるのでしょうか? これもまた前回のくりかえしになりますが，**個体差の正体** に関する生物学的な解釈について説明しておきます．もしこれが何か現実の植物だとすると，観察した個体ごとに体の大きさや年令が違っているとか，個体ごとにもっている遺伝子がちがっているという可能性はあります．これはいかにも個体差らしい要因です．ところが他にも要因はいろいろと考えることができて，たとえばその植物個体が育っている場

9. 8 個中 4 胚珠が結実する個体数は 27.3 と期待されるが観察データでは 6 個体しかいない

10. 結実数 0 個の期待個体数は 0.4 なのに 19 個体，結実数 8 個の期待個体数は 0.4 なのに 20 個体が出現した

11. さらに前回の例で示したように，パラメーターの推定値そのものがおかしくなっている可能性もあります．

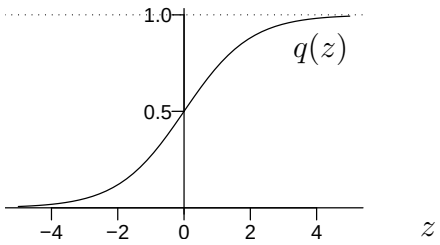
所の明るい・暗い，あるいは土壌中の栄養塩類の多い・少ないで結実確率がちがっているのかもしれません．そうだとすると，これは個体差というよりむしろ場所差みたいなものでしょう．ともあれこういった個体差の原因になりそうなあれこれを**ことごとく観察してデータ化するなんてことは不可能**です．

5. 統計モデルに個体差パラメーターをいれてみる

結実する確率 q が個体によって異なるよう統計モデルを拡張する準備として，結実する確率 $q(z)$ をロジスティック関数 $q(z) = 1/\{1 + \exp(-z)\}$ で表現することにします．

ある個体 i の z を z_i として，

$$q(\beta + r_i) = \frac{1}{1 + \exp(-(\beta + r_i))}$$



というふうに全個体共通の部分 β と個体差 r_i の部分に分割します．このモデルは前回の GLMM 説明で使ったモデルの「切片」β_1 を β にして，葉数依存性の β_2 を無くしてしまったモデルになっています．個性差 r_i ありの二項分布モデルの尤度方程式は

$$L(\beta, \{r_i\} \mid \{y_i\}) = \prod_{i=1}^{100} f(y_i \mid q(\beta + r_i))$$

となります．

個体差なしモデルのときのように，この尤度を最大化するようなパラメータを推定すればよいのでしょうか? 個体差をあらわすパラメータ $\{r_1, r_2, \cdots, r_{100}\}$ は 100 個あり，これに加えて結実する確率のうち全個体共通部分 β を加えると 101 個のパラメータを推定することになります．多少工夫すれば推定すべきパラメータ数を 100 個にできます．いずれにせよ 100 個体の挙動を説明するために 100 個以上のパラメータの推定値を確定しています……こういう「説明」で良いなら統計モデリングなんかはいりません．もっともっと**少ないパラメーターがわかれば特徴づけられる統計モデル**で現象を説明しよう，としているわけです．

6. 階層ベイズモデルで表現する個体差

個体差なしモデルでは観察データにみられるパターンがうまく説明できてるようには見えない，しかし個体差パラメータ $\{r_i\}$ を 100 個も最尤推定するのはいかにもおかしい，という状況を改善するために**階層ベイズモデル**を導入します.

さて，それではこの例題におけるベイズモデルはどう定式化されるのでしょうか? 端的に言うと全 100 個体の個体差 r_i をいちいち最尤推定しない，ということになります．たとえば個体番号 $i = 1$ の架空植物の個体差 r_1 が $-1.2345\cdots$ などと確定できるはずだとは考えないで，r_1 は -3 ぐらいかもしれないし $+0.5$ ぐらいかも，などと**いいかげんなまま放置する**ことにします.

しかしながら，個体差 r_i の確率分布は好き勝手に決めてよし，と許可してしまうとかなり無秩序な推定結果になります．そこで，各 r_i の確率分布は観察データ $\{y_i\}$ と「観察された 100 個体の結実確率には，どこか似ている部分がある」というルールによって制約してしまいたい，つまり「観察データをうまく説明できる範囲で，**個体たちはできるだけ似ている** (r_i がゼロに近い) となるように r_i を決めようね」と，なかなか都合のよいことをもくろんでいるわけです．このように $\{r_i\}$ を制約する役目を与えられた確率分布をベイズ統計学では事前分布とよびます．これに対して，観察データと事前分布で決まる r_i の確率分布は事後分布です.

この個体差 r_i の事前分布は，ここでは簡単のため平均ゼロで標準偏差 σ の正規分布 [12]

$$g_r(r_i \mid \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \frac{-r_i^2}{2\sigma^2},$$

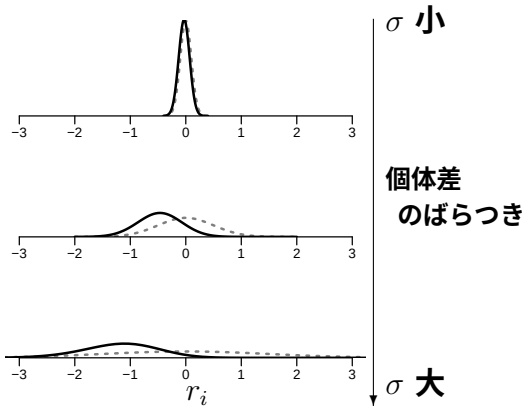
で表現できることにしましょう.

12. これは前回の GLMM 説明で個体差パラメーター r_i の分布を正規分布にしたのと同じです.

観察された個体全体に共通するパラメータ σ は，この植物の個体たちはおたがいどれぐらい似ているかをあらわしていて，たとえば,

- σ がとても小さければ個体差 r_i はどれもゼロちかくになりますから「どの個体もおたがい似ている」
- σ がとても大きければ，r_i は各個体の結実数 y_i にあわせるような値をとる

といった状況が表現できるようになりました．ある個体 i の r_i の事後分布が事前分布 $g_r(r_i \mid \sigma)$ に依存している様子を示します．こ



こで灰色の破線は $\{r_i\}$ の事前分布，黒い実線はある個体 i の r_i の事後分布です．全体のばらつき σ が小さいと r_i はゼロに近く，σ が大きいときには事前分布による制約 [13] が弱くなるのでゼロからずれています．

観察された 100 個体の集団で個体差のばらつきをあらわすパラメータ σ をどう決めるか，が重要な問題になります．しかし，ここではこの問題を先おくりすることにして，ただ単にパラメータ σ もまた何か確率分布 $h(\sigma)$ にしたがう確率変数だ，ということにしてしまいます．これは事前分布のパラメータの事前分布なので**超事前分布** (hyper prior) とよばれています．[14] そしてこんなふうにパラメータを何もかも確率変数にするのであれば，は全個体共通部分をあらわすパラメータ β も事前分布 $g_\beta(\beta)$ にしたがうとしましょう．「σ の超事前分布 $h(\sigma)$ や β の事前分布 $g_\beta(\beta)$ って**どういう確率分布なの?!**」といった疑問なんかも**あとまわし**にします．

めんどうな問題をすべて先おくりにしたまま，ベイズの公式をもちだしてみると

$$p(P \mid D) = \frac{p(D \mid P) \times p(P)}{p(D)}$$

ここで $p(P \mid D)$ は何かデータ (D) のもとで何かパラメーター (P) が得られる確率 (つまりこれが事後分布); $p(D \mid P)$ はその逆でパラメーターを決めたときにデータが得られる確率となるのでこれは尤度 [15] ; $p(P)$ はあるパラメーター P が得られる確率なので事前分布の定義; そして分母の $p(D)$ は「てもとにあるデータ D が得られる確率」というナゾの確率になります．[16] 書きなおしてみると，

$$\text{事後分布} = \frac{\text{尤度} \times \text{事前分布}}{(\text{データが得られる確率})}$$

となり，この関係に架空植物の結実確率を求める例題の統計モデルをあてはめると，観察データ $\{y_i\}$ のもとで

$$p(\beta, \{r_i\}, \sigma \mid \{y_i\}) = \frac{\prod_{i=1}^{100} f(y_i \mid q(\beta + r_i))\ g_\beta(\beta)\ g_r(r_i \mid \sigma)\ h(\sigma)}{\int\int\int (\text{分子} \uparrow \text{そのまま})\ dr_i\ d\sigma\ d\beta}$$

となります．この中で右辺の分母がまったくナゾの確率で決めようがないのですが，ともかくパラメーター β その他に依存してない**ナゾの定数**だということはわかります．分母が定数なので，

$$p(\beta, \{r_i\}, \sigma \mid \{y_i\}) \propto \prod_{i=1}^{100} f(y_i \mid q(\beta + r_i))\ g_\beta(\beta)\ g_r(r_i \mid \sigma)\ h(\sigma)$$

13. ゼロ付近に集まれというしばり,「個性」なんて主張するな，という束縛ですね.

14. 設定したければ超々事前分布でも超々々事前分布でも……

15. ただし，尤度関数の定義なんかでは形式的に $L(P \mid D)$ と書くのがふつうです.

16. 実際のところ，たいていの問題でこれは計算不可能な量です.

すなわち，事後分布の確率密度は尤度 (観察データのもとでの) と事前分布・超事前分布の確率密度の積になっています．個体差パラメータ r_i の事後分布を得るためにその事前分布 $g_r(r_i \mid \sigma)$ が必要で，さらにこの事前分布を決めるために超事前分布 $h(\sigma)$ が必要になる，といった**階層構造**があるので，[17] このような統計モデルは階層ベイズモデルと呼ばれています．

なんだかややこしそうになってきましたが，観測データからこのモデルを特徴づけるパラメーターは推定できるのでしょうか? 階層ベイズモデルのパラメータを推定する方法としては**経験ベイズ法**と **Markov Chain Monte Carlo (MCMC) 法**がよく使われています．

17. ここでいう階層構造なるものが，事前分布は超事前分布に制約されていて，といった nested な関係に限定されていることに注意してください．

7. 経験ベイズ法による最尤推定，すなわち GLMM

まず**経験ベイズ法** (empirical Bayesian method) について説明しましょう．この方法は結局のところ前回に説明した GLMM と同じ推定計算方式，つまり R の `glmmML()` を使うことになります．

前の節で定義した事後分布 $p(\beta, \{r_i\}, \sigma \mid \{y_i\})$ において全個体共通部分のパラメータ β の事前分布 $g_\beta(\beta)$ と個体差のばらつきをあらわす σ の (超) 事前分布 $h(\sigma)$ を「分散がとても大きな一様分布」にしてしまいます．これは言いかえると「β も σ も (観察データにあうように) 好き勝手な値をとっていいよ (ただし σ は標準偏差なので $\sigma > 0$)」と設定していることになります．いっぽうで，各個体の個体差 r_i は平均ゼロかつ標準偏差 σ の正規分布である事前分布 $g_r(r_i \mid \sigma)$ に制約されているので，好き勝手な値をとることはできません．

一様分布の仮定によって $g_\beta(\beta)$ と $h(\sigma)$ が何やら都合よく「定数みたいなもの」に変えらてしまったので，事後分布は

$$\prod_{i=1}^{100} f(y_i \mid q(\beta + r_i))\ g_r(r_i \mid \sigma),$$

に比例する量となり，さらにこの r_i について積分した量は

$$L(\beta, \sigma \mid \{y_i\}) = \prod_{i=1}^{100} \int_{-\infty}^{\infty} f(y_i \mid q(\beta + r_i))\ g_r(r_i, \sigma) dr_i \times (\text{定数})$$

というふうに，観察データ $\{y_i\}$ のもとでのパラメータ β と σ の尤度方程式とみなせます．

あとは尤度 $L(\beta, \sigma \mid \{y_i\})$ を最大化するような $\hat{\beta}$ と $\hat{\sigma}$ を推定すればよいだけです．この数値計算を簡単にすませる抜け道があります．上のような尤度

であらわされる統計モデルは一般化線形混合モデルとまったく同じ形式になります. [18] この種子結実の統計モデルのような (個体差が $\{r_i\}$ だけであるような) 単純な GLMM の場合, 前回と同じく R に glmmML() で β と σ を数値的な最尤推定が可能でしたね.

18. 前回の講義ノートも見てください.

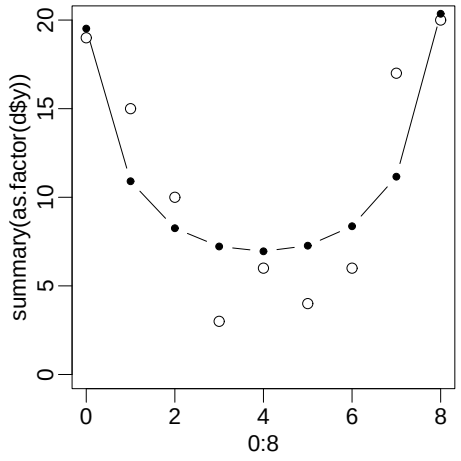
```
> library(glmmML)
> glmmML(cbind(y, 8 - y) ~ 1, family = binomial, cluster = id, data = d)
...(略)...
                coef se(coef)      z Pr(>|z|)
(Intercept) 0.04496   0.3108 0.1447    0.885

Standard deviation in mixing distribution:  2.769
Std. Error:                                 0.2123
...(略)...
```

このように全個体共通部分の $\hat{\beta} = 0.045$ (個体差ゼロの個体の結実確率は $q(\hat{\beta}) \approx 0.51$), そして個体差のばらつき $\hat{\sigma} = 2.77$ などが得られます. [19] これらを使って, 観察データに統計モデルの予測を重ねてみるとこのようになります.

19. なお, この例題における “真の値” は $\beta = 0$, $\sigma = 3$.

```
> library(glmmML)
> source("g.R")
> fit <- glmmML(cbind(y, 8 - y) ~ 1, family = binomial,
+ cluster = id, data = d)
> beta <- fit$coefficients[1]
> sigma <- fit$sigma
> plot(0:8, table(d$y), ylim = c(0, 20))
> lines(0:8, d.gaussian.binom(0:8, 8, mu = beta, sd = sigma) * 100,
+ type = "b", pch = 20, lty = 1)
```



ここで関数 d.gaussian.binom() [20] は

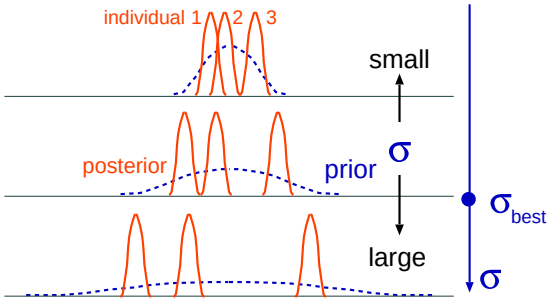
20. これは g.R の中で定義されています.

$$100 \int_{-\infty}^{\infty} f(y \mid q(\hat{\beta} + r)) g_r(r \mid \hat{\sigma})\, dr,$$

というふうに y の事後分布の期待値を計算しています.

結実確率に個体差を加えることで，観察されたパターンをよりよく説明できているような予測が得られました．つまり統計モデルの改善によって,「結実確率 $q(z)$ は 0.5 ぐらい」という全体の平均だけでなく,「$z = \beta + r_i$ とすると，結実確率の個体差 r_i 全体のばらつき σ は 2.8 ぐらい」という個体差に関する知識も新しく獲得できました.

ここで GLMM の推定計算関数 `glmmML()` がやっているのは，下の図でいうと [21]

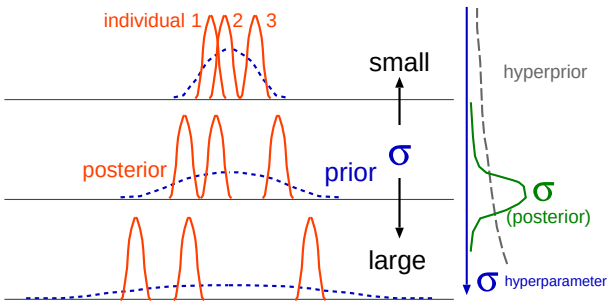


21. この図はかなり「正確さよりわかりやすさ優先」の方針で描かれています.

架空植物の (標本) 集団内の個体差のばらつきをあらわすパラメーター σ を変えると r_i の事後分布の「はなれぐあい」が変わり，あてはまりの良さと「r_i の事前分布の広さ」が変わる，このバランスで事後確率が良くなる何か特定の $\hat{\sigma}$ という値があるはずだから最尤推定しよう，という方針で計算していることになります.

8. Markov chain Monte Carlo 法とは何か?

「何でも最尤推定できるはず」と考える経験ベイズ法に対して,「しかしながらパラメーター数が多くなってきたら**それは無理**でしょう」と考えて，この図のように



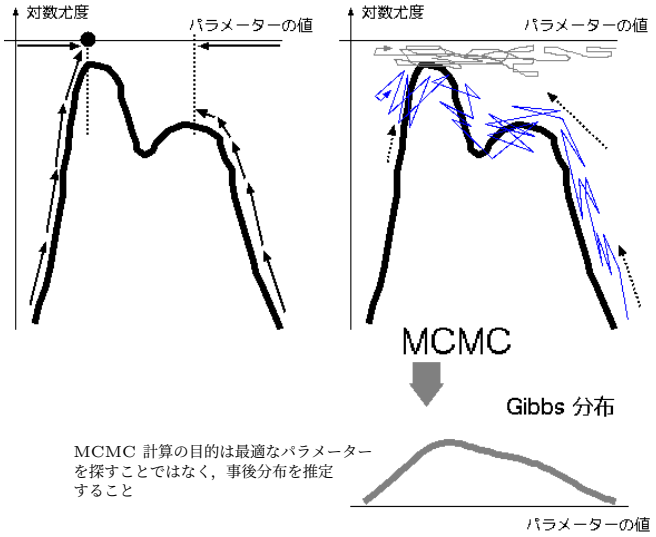
σ の事後分布を推定してやればよい，と考えるのが階層ベイズモデルの **Markov Chain Monte Carlo (MCMC) 法**による推定です.

個体差を考慮したモデルが今回の例題のように簡単なものであれば経験ベイズ法でも十分に解決できます．しかしながら，現実のデータ解析ではもっとあちこちに個体差などが入った統計モデルを構築しなければならない状況が多くなります．個体差 r_i のたぐいがたくさん入った統計モデルでは r_i たちの積分計算は [22] 事実上不可能になり，経験ベイズ法ではパラメータの推定計算ができません．このようなときには MCMC 法によって，パラメータの事後分布をサンプルしていく方法が使われます．

22. 計算量が増大するので

最尤推定法と MCMC 法のちがいをまとめてみますと， [23]

23. 今日はいつもにもましていいかげんな図ばかりだな……



- 最尤推定法はひたすら「対数尤度最大!」の点をめざす
  - それに対して MCMC 計算はできるだけあてはまりの良いところに行くけれど，ときどき悪い方向にも動く
- 最尤推定法は「にせピーク」につかまることが多い
  - MCMC 計算はそのあたりかなりマシ (…… だといいのになぁ，ぐらいの場合もあります)
- 最尤推定法は最尤推定値をもとめることが目的
  - MCMC 計算はふらふらさまよい歩くことでパラメーターの事後分布を「推定」(つまり事後分布からのサンプリング) をする

……といったちがいがあります．

また MCMC 法の中にもいくつもの計算技法があります．代表的なものは次の二つです．

- **Metropolis-Hastings (MH) 法**: もっとも基本的な MCMC 計算法. パラメーターの値をちょっと変えてみる → 新しい値に移動すべき確率を計算 → 乱数によって移動する・しないを決める，という手順を繰りかえします.
- **Gibbs sampling 法**: MH 法より「効率良く」事後分布からのサンプリングをめざす方法．事後分布から直接乱数を生成させる．Gibbs sampling する計算プログラムなどを Gibbs sampler とよぶ.

……MCMC 計算についてはもっと詳しく説明すべき，だとは思うのですが時間がない [24] のでこれだけにしておきます．MCMC 法に関する参考文献としては以下のものをあげておくので，興味がある人は参考にしてください.

24. 説明の時間も…… ああ，そして準備の時間も!

- 伊庭幸人, 『ベイズ統計と統計物理』, 岩波書店, 2003.
- 伊庭幸人・種村正美・大森裕浩・和合肇・佐藤整尚・高橋明彦. 『計算統計 II マルコフ連鎖モンテカルロ法とその周辺』, 岩波書店, 2005.
- 丹後俊郎, 『統計モデル入門』, 2000, 朝倉書店

9. 階層ベイズモデルの MCMC 計算

架空植物の種子結実確率の問題にもどります．階層ベイズモデルによって，パラメーターの事後分布はこのように定義されました.

$$p(\beta, \{r_i\}, \sigma \mid \{y_i\}) \propto \prod_{i=1}^{100} f(y_i \mid q(\beta + r_i))\ g_\beta(\beta)\ g_r(r_i \mid \sigma)\ h(\sigma)$$

この事後分布からのサンプリング，つまりこのモデルで使っているパラメーター $\{\beta, \sigma, r_1, r_2, \cdots, r_3\}$ が「どういう値になりそうか」を MCMC 法によって推定してみましょう.

その前に (今まで放置していた) β の事前分布である $g_\beta(\beta)$ と σ の超事前分布である $h(\sigma)$ を決めねばなりません.「平均的な結実確率」をあらわす β に関しては先験的な情報が何もないので，その事前分布は**無情報事前分布** (non-informative prior) と設定します．たとえば平均ゼロで標準偏差 10 の「すごくひらべったい正規分布」,

$$g_\beta(\beta) = \frac{1}{\sqrt{2\pi 10^2}} \exp \frac{-\beta^2}{2 \times 10^2},$$

なんかを使うことにしましょう.「架空植物たちの個体差のばらつき」こと σ に関しても情報がないので，$h(\sigma)$ はやはり無情報事前分布にします．ただ

し $\sigma > 0$ でなければならないので，正規分布ではなく「すごくひらべったいガンマ分布」を使うことにしましょう．

このように階層ベイズモデルの詳細がきまれば，あとは Gibbs sampler ソフトウェアを使って事後分布を推定するだけです．R 用の使いやすい Gibbs sampler がないので，WinBUGS というソフトウェアを使うことにします……ただしこれを「R の部品」であるかのように使うことにします．まず，この結実確率の階層ベイズモデルを BUGS 言語で coding してやります．

```
model
{
    for (i in 1:N) { # 全サンプル i = 1, 2, ..., 100 について
        Y[i] ~ dbin(q[i], 8) # 二項分布
        logit(q[i]) <- beta + r[i] # 結実確率
    }
    Tau.noninformative <- 1.0E-2 # 分散の逆数
    beta ~ dnorm(0, Tau.noninformative) # 無情報事前分布
    for (i in 1:N) {
        r[i] ~ dnorm(0, tau) # 個体差
    }
    P.gamma <- 1.0E-2 # 無情報ガンマ事前分布のパラメーター
    tau ~ dgamma(P.gamma, P.gamma) # 無情報事前分布
    sigma <- sqrt(1 / tau) # 分散逆数→標準偏差 (output 用)
}
```

今回の例題は比較的簡単なので，BUGS コードも単純なものになりました．もう少し解説してみると，

```
Y[i] ~ dbin(q[i], 8) # 二項分布
logit(q[i]) <- beta + r[i] # 結実確率
```

確率 `q[i]` が `beta` と `r[i]` が決まっているときに，`Y[i]` が二項分布 (確率 `q[i]`，胚珠数 8) から得られる確率を評価しなさい，といった意味になります．

またパラメーター `beta` と `r[i]` については

```
beta ~ dnorm(0, Tau.noninformative) # 無情報事前分布
for (i in 1:N) {
    r[i] ~ dnorm(0, tau) # 個体差
}
```

ここでは右辺で定義されている正規分布 (`dnorm(...)`) から `beta` と `r[i]` の新しい値をランダムサンプリングしなさい，ただし上で定義している「種子数データ (`Y[i]`) へのあてはまりが**あまり**悪くならないように」といったような制約がつけられています．

さて，この階層ベイズモデル定義ファイルだけでは WinBUGS は動いてくれません．観測データやパラメーターの初期値，いったい何回ぐらいサンプリングするのか，といった指示も必要です．これを WinBUGS 上でいちいち指示するのはたいへんなので，R の中でそういったことを準備して「WinBUGS を R の下うけ」として働かせるのがうまいやりかただろうと思います．それを可能にするのが R の R2WinBUGS package です．

ただし R2WinBUGS もあまりデキがよろしくないので，私は R2WinBUGS を**もっと快適につかう関数セット**というのを自分で作ってしました．それを R2WBwrapper.R というファイルにまとめているので，[25] それを読みこんだうえで上の BUGS コードで定義した階層ベイズモデル [26] を WinBUGS に下うけ計算させる R コードはこのようになります．

```
# 関数定義とデータの読みこみ
source("R2WBwrapper.R")
d <- read.csv("data7a.csv")

# 観測データを設定する
set.data("N", nrow(d)) # 標本数 (100)
set.data("Y", d$y) # 結実種子数

# パラメーターの設定
set.param("beta", 0) # 平均結実確率
set.param("r", rnorm(N, 0, 0.1)) # 個体差
set.param("tau", 1) # 分散の逆数
set.param("sigma", NA) # sqrt(tau の逆数)
set.param("q", NA) # 個体ごとの結実確率

# WinBUGS の実行
post.bugs <- call.bugs(n.iter = 700, n.burnin = 100, n.thin = 3)
```

最後の WinBUGS の実行 (そして結果を post.bugs に格納) するところでは計算 step 数を指定しています:

- n.iter = 700: 全体で 700 MCMC step 計算せよ
- n.burnin = 100: ただし最初の 100 step は焼き捨て (burn-in) せよ
- n.thin = 3: 101-700 step の 600 step に関しては，3 step とばし，つまり合計 200 sample を取れ

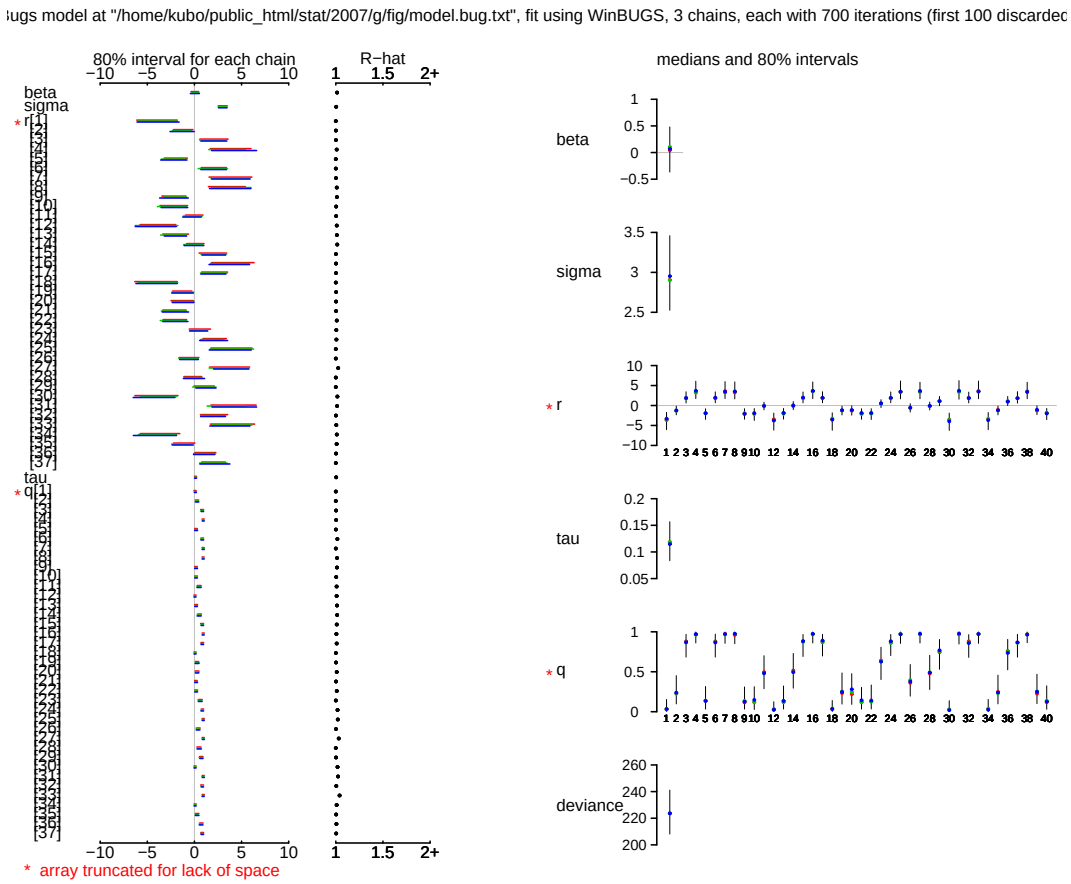
あとは R を起動して source("runbugs.R") とすれば R から WinBUGS が呼びだされて下うけ計算をやってくれます．Gibbs sampling による MCMC 法で得られた推定計算の結果は post.bugs という bugs クラスのデータオブジェクトに格納されています．

25. 例によって講義ページからダウンロードできます．R2WinBUGS package だけでなく coda もインストールする必要があります．

26. これは model.bug.txt というテキストファイルとして保存されているものとします

10. 事後分布として表現される推定結果

結果の概要は plot(post.bugs) で見ることができます.



上の図についてだいたいのところを説明してみますと……

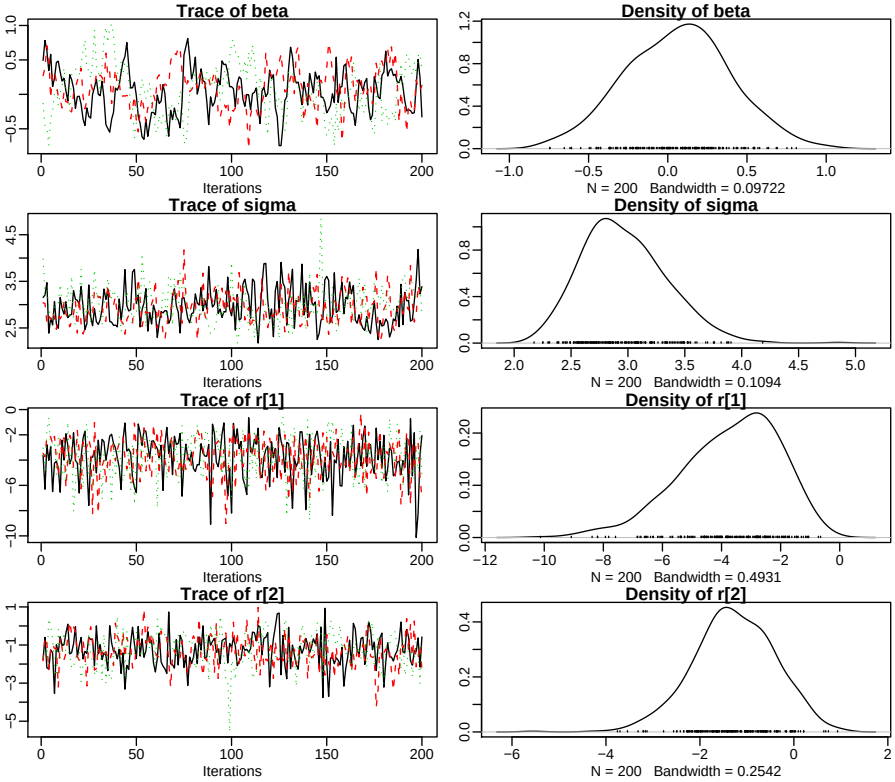
- 図の左側: MCMC 計算の収束ぐあいを示している．パラメーターごとに Gelman-Rubin の $\hat{R}$ 値が示されていて，これが 1 に近いほど収束がよい，つまり MCMC 計算がうまくいってることを示している
- 図の右側: パラメーターの事後分布の平均値 (点) と 80% 信頼区間 (縦のバー)
- 左右どちらの図でもパラメーターが多すぎて全部を示せない場合は途中でうちきっている (* マークがついている)

またひとつひとつのパラメーターが MCMC 法によってサンプリングされている様子や事後分布のカタチなんかも R に作図させることができます.

```
> post.list <- to.list(post.bugs)
> plot(post.list[,1:4,], smooth = F)
```

次の図の左側は MCMC step 数とともに変化しているパラメーターごとの

サンプリングの様子，右側は事後分布の近似的な確率密度分布を示しています．



さてさて，ここで今回の統計モデルについて復習してみると，結実確率 $q_i = q(\beta + r_i)$ はこんな式で

$$q(\beta + r_i) = \frac{1}{1 + \exp(-(\beta + r_i))}$$

さらに「個体差」 r_i はこんな正規分布

$$g_r(r_i \mid \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \frac{-r_i^2}{2\sigma^2},$$

にしたがう，としていました．「個体差」のばらつきが σ です．上の事後分布の図をみると，β は中央値がゼロぐらいの確率分布，σ は中央値が 3 ぐらいの確率分布となっています．これらの中央値はこの架空例題をつくるときに与えた値 ($\beta = 0, \sigma = 3$) になっています．[27] つまりうまく推定できているわけですが，ベイズの良いところは「けっきょくパラメーターの推定値って**有限個のデータ**から推定したものにすぎないのであって，推定のふらつきがあるはず」という推定の不確定性の部分を事後分布としてうまく表現できている，という点にあります．上の図で示されている確率分布 (確率密度分布) は「パラメーターがこのへんにある確率 (確率密度)」をあらわしています．

27. また個体ごとの r_i の事後分布も上の図で一部 (`r[1]` と `r[2]`) が示されています．

パラメーターの事後分布に関する table はこのように出力します．ここから値をとりだせば，この階層ベイズモデルの予測が観測データにあてはまってるかどうか，といったことも図示できます．

```
> print(post.bugs, digits.summary = 2)
... (略) ...
 3 chains, each with 700 iterations (first 100 discarded), n.thin = 3
 n.sims = 600 iterations saved
           mean    sd   2.5%    25%    50%    75%  97.5% Rhat n.eff
beta       0.07  0.33  -0.59  -0.18   0.08   0.28   0.71 1.01   390
sigma      2.97  0.37   2.36   2.70   2.92   3.20   3.76 1.00   450
r[1]      -3.72  1.67  -7.55  -4.76  -3.51  -2.50  -1.12 1.00   600
r[2]      -1.28  0.87  -3.13  -1.81  -1.29  -0.65   0.26 1.00   600
... (略) ...
r[99]      1.97  1.13  -0.02   1.19   1.88   2.62   4.45 1.00   600
r[100]    -3.86  1.75  -7.76  -4.95  -3.64  -2.54  -1.23 1.00   400
tau        0.12  0.03   0.07   0.10   0.12   0.14   0.18 1.00   450
q[1]       0.06  0.07   0.00   0.01   0.03   0.08   0.26 1.00   600
q[2]       0.26  0.14   0.05   0.16   0.23   0.34   0.58 1.00   600
... (略) ...
q[99]      0.85  0.11   0.56   0.78   0.87   0.93   0.99 1.00   600
q[100]     0.06  0.07   0.00   0.01   0.03   0.08   0.25 1.01   290
deviance 223.98 13.10 200.58 214.30 223.55 233.10 250.50 1.00   600

For each parameter, n.eff is a crude measure of effective sample size,
and Rhat is the potential scale reduction factor (at convergence, Rhat=1).

DIC info (using the rule, pD = Dbar-Dhat)
pD = 79.0 and DIC = 302.9
DIC is an estimate of expected predictive error (lower deviance is better).
```

これらの推定結果をくみあわせることで，さきほどの GLMM 推定結果で示したような「ゆがめられてしまった二項分布」の作図ができるだけでなく，「そのばらつき」なんかもわりと簡単に図示できます．

えー，最後のあたりはカケ足な説明になってしまいましたが……今回説明できるのはこんなところでしょうか? さてさて……いろいろと説明できていない部分もありますが，このたびの「生態学の統計モデリング」講義はこのあたりでいったん終了することにしましょう．皆さん，参加してくださって，どうもありがとうございます．