

## 11. 6 Multiple Groups

集団を固定された層(オス/メスなど)にグループ分けして研究することはよくある。性別のように層間で移動がない場合もあれば、年齢のように必ず(deterministicに)推移していくものもある。グループメンバシップはカテゴリカル共変量とみなすことができる。捕獲した動物はすべてこのカテゴリカル共変量を測定できるが、捕獲したことのない動物に対してはグループ分類ができないので、モデル化する必要がある。

幸運なことに、捕獲したことのない動物のモデリングはデータ拡大法(Data-augmentation)の単純な拡張ですむ。必要なことは、捕獲したことのない  $M - \mu$  (\*個体群の個体総数 - 一度でも捕獲された個体の合計) の個体データを NA として、グループ変数を導入するだけである。グループはカテゴリカルランダム変数としてモデルに組み込まれる。その際、カテゴリーの数はグループの数に等しくなる。

Nichols(2005) study on male and female meadow voles  
two-group CMSAモデルを適用 (Panel 11.3を元にPanel 11.5の変更を加えた)  
推定結果はFig. 11.5参照

Panel 11.5

```
for (i in 1:M){
  w[i] ~ dbern(psi)
  b[i] ~ dcat(beta[]) I(birthCT[i])
  grp[i] ~ dcat(pi_grps[b[i],])
}

for (i in 1:k){
  pi_grps[i, 1:ngroups] ~ ddirch(alpha_grp[i,])
}
```

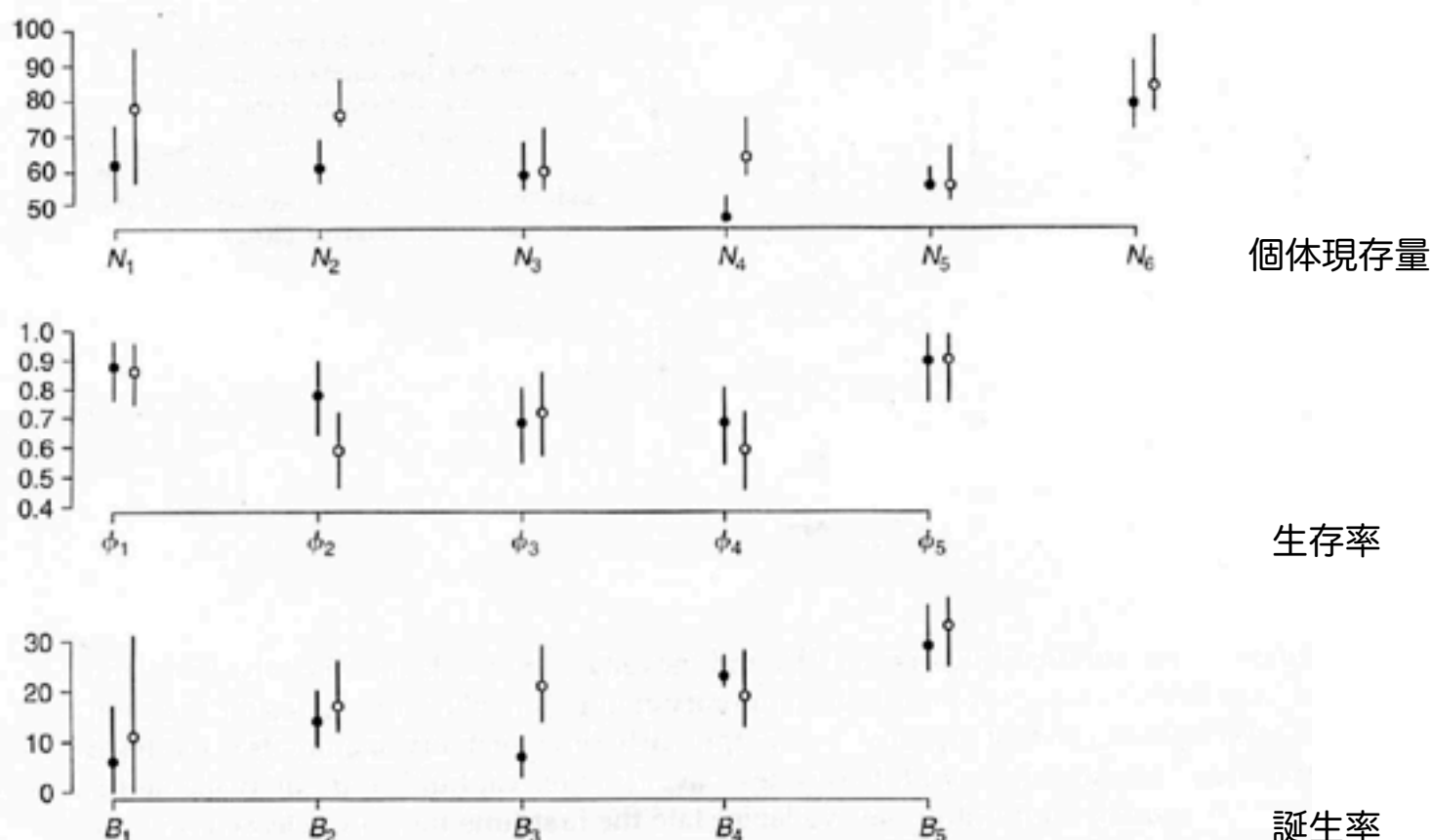


FIGURE 11.5 Posterior density summaries for abundance ( $N$ ), births ( $B$ ), and survival rates from fitting a sex-specific CMSA model to the male and female meadow vole data of Nichols (2005). The closed circles are for males and the open circles for females.

## 11.7 Robust Design

Pollock's robust design (Pollock, 1982)

CMSAモデルから始めてみる。

ロバストデザインでは、集団はあるサンプリング期間で閉じているが、その他の期間には開いているとする。閉じてる期間には $M_t$ や $M_h$ といったモデルで現存量や生存率を推定するのに使えるかもしれない。

CMSAモデルを修正し、生存率 $\phi_j$ を1、移入率 $\beta_j$ を0に制限すると、 $M_t$ モデルを各閉期間に適用したのと同じことになる。開期間では生存率と誕生率に単純な時間依存がある。(?)

BUGSコードにするには、Panel 11.3に下記を追加して修正するのが効率的。

K : 第一期間 number of primary periods

k : サンプリング総数

prim : k次元のベクトル k-dimensional vector

要素jはサンプリング機会が属する第一期間の指標

Panel 11.6

```
model{
  for ( i in 1 : M){
    w[i] ~ dbern (psi)
    b[i] ~ dcat (beta[] ) I (,birthCT[i] )
  }
  Lambda<- sum( z[ ] )
  for ( i in 1 : M){
    for ( j in 1 : AvailUntil [i] ) {
      aCap [ i, j ] <- a[ i, Prim[j] ]
      pCap [ i, j ] <- aCap[ i, j ] * p[ j ]
      X[ i, j ] ~ dbern (pcap[ i, j ] )
    }
    a[ i, 1 ] <- w[i] * (1 - notyetborn[ i, 1 ] )
    avail [ i, 1 ] <- a[ i, 1 ] * step(Prim[AvailUntil [i] ] - 1 )
    for ( j in 2 : K ) {
      a[ i, j ] <- w[i] * (1 - notyetborn [i, j] ) * notyetdead[ i, j ]
      avail[ i, j ] <- a[ i, j ] * step( Prim[ AvailUntil [i] ] - j )
      asuse[ i, j ] <- notyetdead [ i, j-1 ] + 1
      notyetdead [ i, j ] ~ dbern ( sv [asuse [i, j], i, j-1 ] )
    }
    for ( j in 1 : K-1 ) {
      sv[ 1, i, j ] <- 0
      sv[ 2, i, j ] <- notyetborn[ i, j ] + (1 - notyetborn [ i, j ] ) * S[ j ]
    }
  }
  # prior distribution for S and p
  for ( j in 1 : K ) {
    p[ j ] ~ dbeta(1, 1)
  }
  for ( j in 1 : K-1 ) {
    S[ j ] ~ dbeta(1, 1)
  }

  beta[1 : K] ~ ddirch(alpha[ ])
  psi ~ dbeta(1, 1)
}
```

### 11.7.1 Example: Meadow Voles

6ヶ月おきの6回の第一期間primary periods(K=6)。  
 それぞれの第一期間に5回ずつ第二機会(=サンプリング機会?)がある。従ってk=30。  
 2回目の第一期間の最後2日でトラップをアライグマに荒らされたため、正確な捕獲数が分からなくなった。  
 そこで Williams et al. (2002) では最初3日間だけのデータを使うことにした。  
 従来の閉期間捕獲モデルを適用するにあたって、各期間の不均一性をモデルに取り入れなくてはならないと彼らは気付いた。

今回の解析は下記のようにする。

$$\text{logit}(p_{ij}) = \gamma_i + \varepsilon_i$$

(著者らは p (primary + individual) と表記)

pはprimary, individual毎に異なっており、両者は線形加算とする。

$\varepsilon_i$  は平均0、分散 $\sigma^2$ の正規分布に従う個体毎のランダム変数。

ただし、同一primary内では個体毎捕獲率は一定。

Panel 11.7

```

for (i in 1:M){
  eps[i] ~ dnorm(0, tau)
  for (j in 1:k){
    logit(p[i, j]) <- gam[Prim[j]] + eps[i]
  }
}
for (j in 1:K){
  gam[j] ~ dnorm(0, 0.01)
}
spd ~ dunif(0, 10)
tau <- 1 / pow(sdp, 2)
    
```

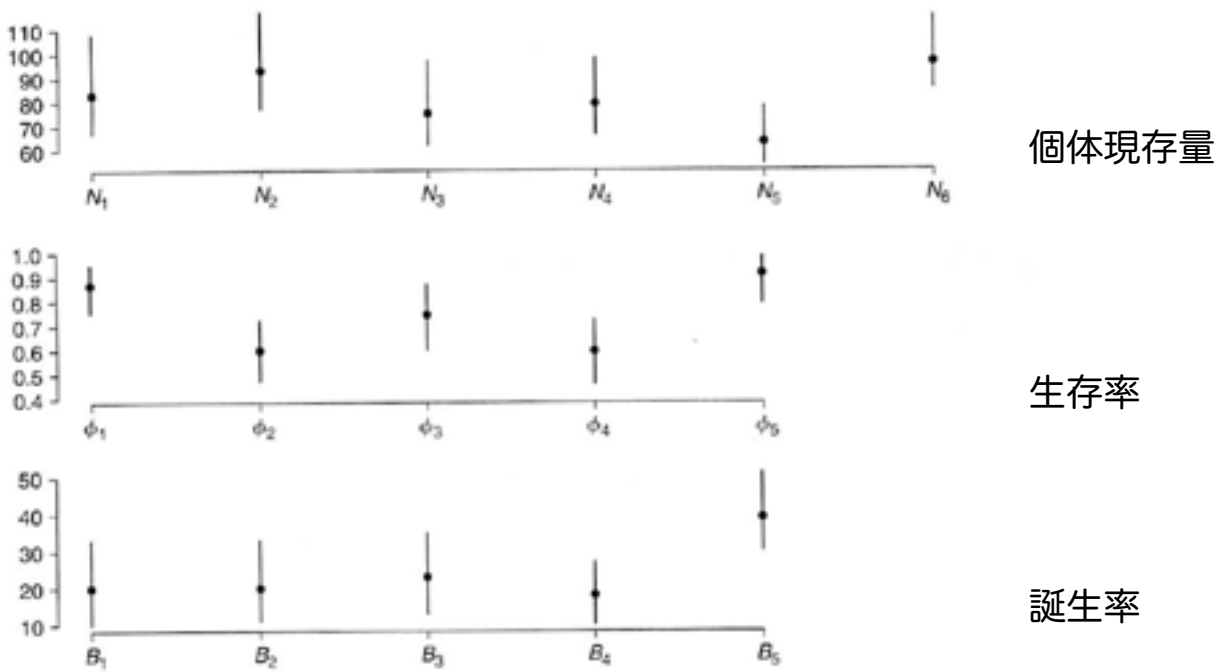


FIGURE 11.6 Posterior summaries for the parameters  $N$  (abundance),  $S$  (survival), and  $B$  (births) for the robust

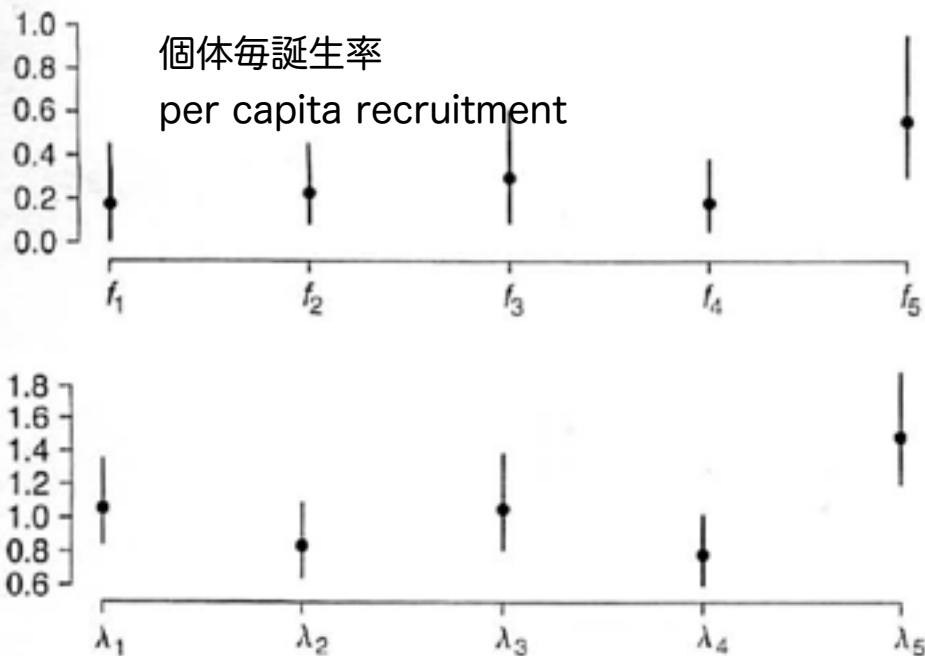


FIGURE 11.7 Posterior summaries for the parameters  $f$  (per capita recruitment) and  $\lambda$  (population growth rate) for the robust design model  $p(\text{primary}+\text{individual})$  in which there are additive effects of primary occasion and individual on  $\text{logit}(p_{ij})$  fitted to the meadow vole data of Williams et al. (2002).

## 11.7.2 Temporary Emigration

固定トラップを用いたマーク再捕獲研究の特徴は、動物が移動する区域はサンプリング区域と正確には一致していないかもしれないということである。

先に述べたモデルは一時的に捕獲不能になる状態に適用でき、動物が死亡した場合と単に移動しただけの場合を分けることができた。

Kendall and Nichols (1995) と Kendall et al. (1997)は一時的移住をロバストデザインにどのように組み込めばよいか示しているが、彼らの尤度は最初の捕獲だけにしか適用できない。

このアプローチでは捕獲可能かどうかはfirst-order Markov chainによりモデリングされ、捕獲可能か捕獲不可能かの2つの状態をとる。

全ての動物は最初に捕獲され、放獣される時捕獲可能の状態から始まり、捕獲される時はこの状態であるとされる。サンプリング  $j$  と  $j+1$  の間に起こる状態の遷移は遷移マトリクス  $\psi_j$  に従って決まる。

$$\psi_j = \begin{pmatrix} 1 - r_j'' & r_j'' \\ 1 - r_j' & r_j' \end{pmatrix}$$

		サンプリング	
		$j-1$	$j$
$r_j''$	捕獲可能	捕獲可能	捕獲不可能
$r_j'$	捕獲不可能	捕獲不可能	捕獲不可能
$1 - r_j''$	捕獲不可能	捕獲不可能	捕獲可能
$1 - r_j'$	捕獲可能	捕獲可能	捕獲可能

ロバストデザインに初めての捕獲というモデルを含めるために、捕獲可能と捕獲不可能の2つの状態を割り当てるパラメータを導入する必要がある。

$\delta_i$ : 動物が  $j-1$  と  $j$  の間に生まれ、 $j$  で捕獲可能になる確率

$\delta_i$ を導入することで、partially observed Markov chainで捕獲可能モデルを記述できるようになる。

$Z_{ij}$ :  $i$  個体がサンプル  $j$  の時点で生きており、捕獲可能であれば状態1、捕獲不可能であれば状態2とする。

$b_i$ : 生誕(生存)期間 birth interval

$d_i$ : 死亡期間 death interval

個体  $i$  の状態は  $\{Z_{i,b_i}, \dots, Z_{i,d_i}\}$  と書き表せ、その確率は、

$$\Pr(Z_{i,b_i} = s | b_i) = \begin{cases} \delta_{b_i} & s=1 \\ 1 - \delta_{b_i} & s=2 \end{cases}$$

$$\Pr(Z_{i,j} = s | Z_{i,j-1} = r, b_i, d_i) = \begin{cases} 1 - r_j'' & r=1, s=1 \\ r_j'' & r=1, s=2 \\ 1 - r_j' & r=2, s=1 \\ r_j' & r=2, s=2 \end{cases} \quad \text{for } j = b_i + 1, \dots, d_i$$

TABLE 11.6 Some Markov chains describing animal availability for capture, and their associated probabilities, conditional on time of birth and death, for a study with four primary periods. State 1 = "available," state 2 = "unavailable," and a value of "NA" is used to represent states that are unassigned, either because the animal is not yet born or because it has died.

Chain	$b_i$	$d_i$	$\Pr(Z_{i,j}   b_i, d_i)$
(1,1,2,1)	1	4	$\delta_1(1 - \gamma_1'')\gamma_2''(1 - \gamma_3')$
(2,1,2,2)	1	4	$(1 - \delta_1)(1 - \gamma_1')\gamma_2''\gamma_3'$
{NA,1,2,1}	2	4	$1 \times \delta_2\gamma_2''(1 - \gamma_3')$
{NA,2,2,2}	2	4	$1 \times (1 - \delta_2)\gamma_2'\gamma_3'$
{NA,1,1,NA}	2	3	$1 \times \delta_2(1 - \gamma_2'') \times 1$
{NA,2,2,NA}	2	3	$1 \times (1 - \delta_2)\gamma_2' \times 1$

## 11.7.2 Example : Meadow Vole Robust Design with Heterogeneity and Temporary

このモデルをmeadow voleのデータに適用する。その際、パラメータ $\delta$ 、 $r''$ 、 $r'$ は時間によらず一定とする。最尤法を用いてこのモデルを当てはめる時はたいてい何らかの時間に対する制約が必要である(Kendall et al., 1997)。一緒に $p$ の不均一性を $p(\text{primary} + \text{heterogeneity})$ の形で盛り込む。

捕獲不可能な状態の時動物は見られないから、 $\Lambda$ や現存量 $N_j$ 、誕生数 $B_j$ なども含め、全てのパラメータが同定可能なわけではない。しかし捕獲の可能性のある集団( $N'$ )について一部は推定できる。また、誕生率パラメータは同定可能であるから、捕獲可能/不可能の状態と同じように誕生と生存の率も有効な推定ができる。

結果はFig.11.8

サンプル  $j$  において捕獲可能な個体数 $N'_j$ 、 $j$  から  $j+1$ の期間の生存率。  $N'$ がFig.11.6の $N$ より小さいが、これは $N$ が集団内の全ての個体数を推定しているのに対し、 $N'$ は捕獲可能な個体だけを推定しているためである。

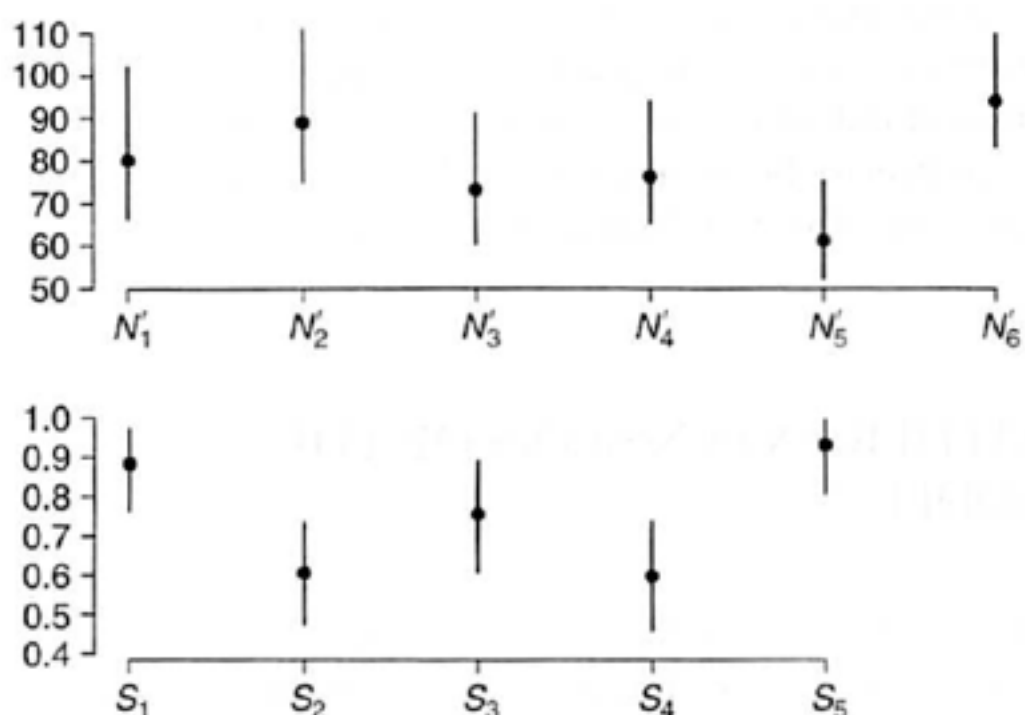


FIGURE 11.8 Posterior summaries for the parameters  $N'$  (abundance of animals available for capture) and  $S$  (survival) for the temporary emigration robust design and meadow vole data of Williams *et al.* (2002) with the model  $p(\text{primary} + \text{individual})$  for capture probabilities in which there are additive effects of primary occasion and individual on  $\text{logit}(p_{ij})$ .

$r''$ 、 $r'$ 、 $\sigma_p$ の結果はFig.11.9

$r''$ の事後分布は0.10以下に偏っているのは、捕獲可能な状態から捕獲不可能な状態へ移るのがほとんどないことを示唆している。 $r'$ の事後分布はデータがこのパラメータに対しあまり情報を持っていないことを示唆している。そもそも捕獲危険下にある集団を移出する事例が少ないので驚くことではないが、もしデータが十分にあったとしても $r'$ の推定は不正確になりがちである。なぜならある個体がサンプリング  $j$  で生きていると確認できたとして、サンプリング  $j-1$  では不可能であったと正確に知るのとは不可能であるからである。

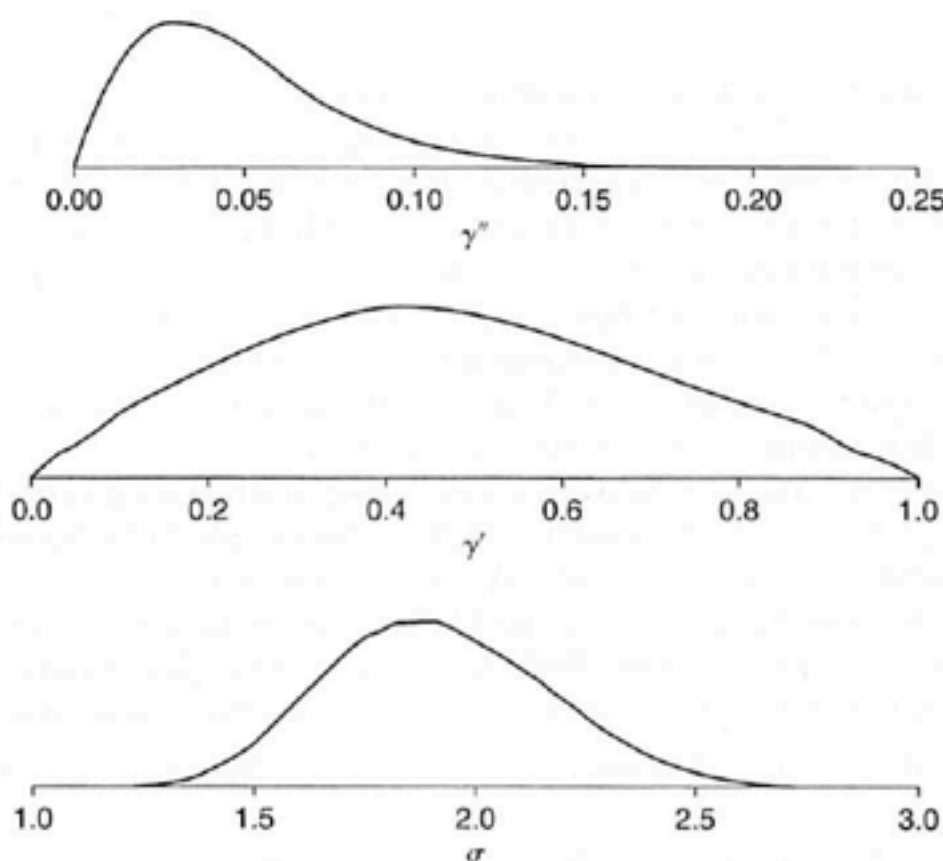


FIGURE 11.9 Posterior summaries for the parameters  $r''$  (probability of moving from the available state),  $r'$  (probability of remaining in the unavailable state), and  $\sigma_p$  (standard deviation of  $\text{logit}(p_{ij})$ ) for the temporary emigration robust design and meadow vole data of Williams *et al.* (2002) with the model  $p(\text{primary} + \text{individual})$  for capture probabilities in which there are additive effects of primary occasion and individual on  $\text{logit}(p_{ij})$ .



## 11.8 Multistate models and other extensions of the CJS model

最後の例では動物が捕獲可能か捕獲不可能かの2つの状態のうちどちらかをとる例を扱った。こういったモデリングは動物集団の研究ではしばしば必要となる。例えば地理的场所、繁殖状態、健康状態、体サイズクラスなどである。これらの状態はそのものが興味の対象かもしれないし、統計量demographic parametersや検出率detection rateを説明するものであるかもしれない。

最も広く使われるマルチステートモデルは、CJSモデルの拡張版であるArnason-Schwarz model である。このモデルの重要な2つの特徴は、(i)初めての捕獲を条件とする(ii)状態は各捕獲時に観察できる。しかしながら、捕獲していない動物の状態は観測できないので、状態は部分的に観測されたカテゴリカル共変量とみなしうる。捕獲不可能な状態をロバストデザインに導入したKendall et al.(1997)のモデルは2つの部分的に観測される状態のマルチステートモデルとみなせる。

Gibbs samplingを開発したDupuisがArnason-Schwarzモデルを初めてベイズで実装したなど、発展の歴史と最近の様子が続く...

## 11.9 Afterword

censoring models, known-fate models, そして完全には発見できない状況でのopen population mark-recapture models、これらはどれも我々の手に入るデータが我々が望むデータ(data we wish we had) そのものでないという点を共有している。ここで提供した解析はCDL'sに基づいていた。その利点は二つある。一つはモデルを作成するのにより直感的で効率的なプログラムコードを書けることである。特にODL'sに基づいた尤度法と比較してGibbs samplingを使ったMCMCは一般にとても簡単である。

第2にCDL'sの利用は生物学的興味のある側面の推定に焦点をあてている。我々が望むデータ(data we wish we had)の多くは第一の興味の統計現象や、推定したいものを記述していたりする。ベイズ法でCDL'sを利用することはこれらの量に対し予測値を与える。

これまでの書誌に登場してきた尤度関数はほとんどすべてODL'sであり、誕生、死亡、捕獲過程といった事象が絡み合ってモデルを理解しがたいものにしてきた。また、マーク再捕獲モデルの標準的パラメタライゼーションの多く、特にJSモデルのような誕生と現存量のモデルで、階層的拡張が容易でなかった。隠れた変数をモデリングすることを可能とするのはこういった階層的拡張である。この追加構造はベイズ解析で簡単に実装できる。

CDLを誕生、グループによる違い、複数の階層、時間とともに変わる共変量、ロバストデザインといったものを含んだCJSモデル基本モデルに拡張することは比較的単純であることを示してきた。この章では議論していないが、CDLにresightingやdead recoveriesといったものから供給される補助データを含めるようにするのもまた簡単である。