

## Chapter 7: Hierarchical Modeling 階層モデル

### 7.2 階層モデル

複数のパラメータを扱う場合に、それらが互いに独立でない(dependent)場合がある。この章以降で扱うものには次の3つがある。

#### 1. Simultaneous estimation of hospital mortality rates

##### 病院の死亡率を同時に推定する

例えば、94カ所の病院での心臓移植手術の死亡率で、各病院の真の死亡率 $\lambda_i$ のとき、94全ての死亡率 $\lambda_1, \dots, \lambda_{94}$ を同時に推定したいとする。真の推定値は同様な値をとると考えるのが普通で、これらのパラメータ間の関係は互いに独立でない。つまり、ある病院の真の死亡率が既知であれば他の病院もそれに近い値だろうと考える根拠となる。

#### 2. Estimating college grade point averages (GPA)

##### 成績の平均値を推定する

10章の例で、大学の入試係では高校ランクと個人の成績別に、新入生の成績(GPA)平均値を集める。その目的は大学に入ってからからの学生の成績を予測するため。集団の平均GPAは高校のランクとテストの成績による単純な線形で代表される。

#### 3. Estimating career trajectories

##### スポーツ選手の年齢とパフォーマンスの関係

11章の例で、スポーツ選手の年齢とパフォーマンスの関係を扱う。9人の野球選手と年齢によるバッティング成績にそれぞれモデルをフィッティングした図7.1で注目したいのは、カーブの形が似ていること。一般に20代後半~30代前半にピークがありその後下がる。つまり、1.と同様、選手同士でカーブが似ていて事前分布は独立でない(dependence)ということ。

このようにパラメータが沢山ある場合、階層的に事前分布を構築するのが普通である。観察値はパラメータで規定される分布に従い、パラメータは超事前分布を規定するハイパーパラメータに規定される分布を持つ。

データ分布は  $y \sim f(y|\theta)$

事前ベクトル $\theta$ は未知のハイパーパラメータ $\lambda$ によって割り当てられる。

$$\theta \sim g_1(\theta|\lambda)$$

ハイパーパラメータベクトル $\lambda$ はある分布を持つ。

$$\lambda \sim g_2(\lambda)$$

一般的に階層的事前分布ではパラメータセット $\theta=(\theta_1, \dots, \theta_k)$ の要素は入替可能(exchangeability)と考える。つまり、 $\theta_1, \dots, \theta_k$ は $g_1(\theta|\lambda)$ からのランダムサンプルとする。そしてハイパーパラメータベクトル $\lambda$ は次のステージの既知のpriorとなる。

$$\lambda \sim g_2(\lambda)$$

この階層的priorを死亡率の例と11章のスポーツ選手の例に応用。

### 7.3 Individual and Combined Estimates

#### 個別あるいは同時推定

3章の心臓移植の死亡率について考える。94のi番目の病院で心臓移植後30日以内の死亡者数 $y_i$ とその病院での手術数(exposure)  $e_i$  (期待値 $e$ ) のデータhearttransplantsを使用。

```
>data(hearttransplants)
>attach(hearttransplants)
> head(hearttransplants,3)
  e y
1 532 0
2 584 0
3 672 2
```

ここで標準化モデルでは死亡者数 $y_i$ は平均 $e_i \lambda_i$ のポアソン分布に従い、手術数( $e$ )に対する死亡率を推定することになる。 $y_i/e_i$ 比は $i$ 病院の死亡率となる。図7.2は $\log(e)$ に対する $y/e$ をプロットしたもので、数字は死亡者数。

```
plot(log(e), y/e, pch=as.character(y))
```

推定した死亡率は違いが大きく、とくに手術数が小さいとその差が大きいに死亡者0が多くなっている。

そこで各病院の真の死亡率 $\{\lambda_i\}$ を一度に推定したい。

まず、個々の死亡率を使う方法。

<第1>

$$y_1/e_1, \dots, y_{94}/e_{94}$$

ただ、手術数 $e$ 少ない病院では良い推定値とはならない。図7.2でわかる通り移植死亡者が0の病院では $y_i/e_i = 0$ は過小評価だし、手術数が少ない病院ではブレが大きい。

これではよい推定値が得られないことから、全てのデータからよりよい推定を得てみる。ここで、真の死亡率が全病院で等しいとする。

$$\lambda_1 = \dots = \lambda_{94}$$

平均が等しい(equal-means)ポアソンモデルのもとでの $i$ 病院での死亡率。

<第2>

$$\sum_{j=1}^{94} y_j / \sum_{j=1}^{94} e_j$$

となるが、全病院で死亡率が等しくなければならない。

そこで、<第3>の推定値

$$(1-\lambda) y_i/e_i + \lambda \sum_{j=1}^{94} y_j / \sum_{j=1}^{94} e_j$$

パラメータ $0 < \lambda < 1$ によって $y_i/e_i$ を $\sum_{j=1}^{94} y_j / \sum_{j=1}^{94} e_j$ に近づけたとなる。これは、真の死亡率による入替事前モデル(exchangeable prior model)から導かれる。

## 7.4 Equal Mortality Rates?

### 死亡率は等しいのか？

exchangeable modelを考える前に、死亡率が等しいときのモデルのフィッティングとその検証を試みる。 $y_i$ がポアソン分布( $e_i \lambda$ ),  $i=1, \dots, 94$ に従い、等しい死亡率 $\lambda$ による無情報標準事前分布,

$$g(\lambda) \propto 1/\lambda$$

事後分布は

$$g(\lambda | \text{data}) \propto \frac{1}{\lambda} \prod_{j=1}^{94} [\lambda^{y_j} \exp(-e_j \lambda)] \\ = \lambda^{\sum_{j=1}^{94} y_j - 1} \exp(-\sum_{j=1}^{94} e_j \lambda)$$

これは、パラメータ $\sum_{j=1}^{94} y_j$ と $\sum_{j=1}^{94} e_j$ に従うガンマ密度関数である。

ここで、

```
> sum(y)
```

```
[1] 277
```

```
> sum(e)
```

```
[1] 294681
```

なので、事後分布の死亡率は $\Gamma(277, 294681)$ となる。

ベイズ法によるモデルの適合性は事後予測分布(posterior predictive distribution)に基づいて行う。 $y_i^*$ を $i$ 病院の将来での手術数 $e_i$ での死亡者数とする。 $y_i^*$ は真の値 $\lambda$ での平均 $e_i \lambda$ のポアソン分布に従う。 $i$ 番目の真の値は事後分布 $g(\lambda | y)$ に含まれる。無条件下での $y_i^*$ , 事後予測分布は

$$f(y_i^* | e_i, y) = \int f_P(y_i^* | e_i \lambda) g(\lambda | y) d\lambda,$$

$f_P(y | \lambda)$ は平均 $\lambda$ のポアソン抽出密度。事後予測分布はフィットモデルに基づいた将来予測の尤度となる。例えば、密度 $f(y_i^* | e_i, y)$ は $e_i$ ある病院の将来の死亡数(?)となる。実際の観察値 $y_i$ が予測分布の真ん中(middle)にあれば、モデルフィットの整合性があると言える。一方、 $y_i$ が $f(y_i^* | e_i, y)$ の極端なテイルにある時は、このモデルは適していないと判断する。

事後予測分布を図示する。病院94の死亡者  $y_{94}=17$ は極端に多いのか？1000個の値を事後予測分布  $y^*_{94}$ からシミュレートする。

```
まず、事後分布から1000回  $\lambda$  を抽出  
lambda=rgamma(1000, shape=277,  
rate=294681)  
これを用いて平均  $e_{94}\lambda$  のポアソン分布から  
 $y^*_{94}$ を導く  
ys94=rpois(1000, e[94]*lambda)
```

事後予測分布Fig. 7.3を書く。太線が94病院の死亡者数  $y_{94}=17$

```
hist(ys94, breaks=seq(0.5, 26.5, by=1))  
lines(c(y[94], y[94]), c(0,120),lwd=3)
```

ここで、観察値  $y_i$ はテイルの部分にあるので、このモデルは当てはまらない。そして、この病院は死亡率一定とすると死亡者数が多いことになる。

全ての病院での観察値の適合性を事後予測分布からチェックすることができる。各分布において将来の予測値  $y^*_i$ が観察値  $y_i$ に対して極端な値をとる確率

$$\min\{P(y^*_i \leq y_i), P(y^*_i \geq y_i)\}$$

全ての観測値について範囲内にある確率ベクトル  $pout$ を求めて、

```
pout=0*y  
lambda=rgamma(1000, shape=277,  
rate=294681)  
for (i in 1:94){  
  ysi=rpois(1000, e[i]*lambda)  
  pleft=sum(ysi<=y[i])/1000  
  pright=sum(ysi>=y[i])/1000  
  pout[i]=min(pleft, pright)  
}
```

死亡率図7.4を書く。

```
plot(log(e), pout, ylab="Prob(extreme)")
```

ここで、15個は0.1よりも低い確率で、死亡率一定モデルは94の病院の死亡率の分布には適していない。そこで、死亡率が異なるモデ

ルを次に説明するexchangeable modelで推定する。

## 7.5 Modeling a Prior Belief of Exchangeability

### 入替できるパラメータによる事前予測モデル

事前予測の最初の段階で真の死亡率  $\lambda_1, \dots, \lambda_{94}$  はガンマ分布  $\text{gamma}(\alpha, \alpha/\mu)$  のランダムサンプルとした。

$$g(\lambda|\alpha, \mu) = \frac{(\alpha/\mu)^\alpha \lambda^{\alpha-1} \exp(-\alpha\lambda/\mu)}{\Gamma(\alpha)}, \lambda > 0.$$

$\lambda$  の平均  $\mu$ , 分散  $\mu^2/\alpha$ 。次に、ハイパーパラメータ  $\mu$  と  $\alpha$  を独立に推定する。 $\mu$  はガンマ分布  $\text{gamma}(a,b)$  で密度  $\mu^{a-1} \exp(-b\mu)$ ,  $\alpha$  は密度  $g(\alpha)$ 。

この事前分布は真の死亡率と正の相関をとっている。例えばハイパーパラメータ  $\mu$  を  $\text{gamma}(10,10)$  からとり  $\alpha$  を定数  $\alpha_0$  とする。

(これは  $g(\alpha)$  の  $\alpha_0$  を1とするのと同じ) この事前分布から下記の方法で  $(\lambda_1, \lambda_2)$  をシミュレートする。

- $\mu$  を  $\text{gamma}(a,b)$  より、 $\alpha$  を事前分布  $g(\alpha)$  より導く
- 各ペア  $(\mu, \alpha)$  について  $\lambda_1, \lambda_2$  を  $\text{gamma}(\alpha, \alpha/\mu)$  から導く

この事前分布から  $(\lambda_1, \lambda_2)$  を500回シミュレーションしたものが図7.5。 $\alpha$  は5. 20. 80, 400,  $\mu$  は  $\text{gamma}(10, 10)$  分布で  $\lambda_1$  と  $\lambda_2$  の中心が1になっている。ハイパーパラメータ  $\alpha$  は精度のパラメータ (a precision parameter) で、 $\alpha$  が無限大になると  $\lambda_1 = \dots = \lambda_{94}$ 。

```
par(mfrow = c(2,2))  
m=500  
alphas = c(5,20,80,400)  
for (j in 1:4){  
  mu=rgamma(m, shape = 10, rate = 10)  
  lambda1 = rgamma(m, shape=alphas[j],  
rate =alphas[j]/mu)  
  lambda2 = rgamma(m, shape=alphas[j],  
rate=alphas[j]/mu)  
plot(lambda1, lambda2)
```

```
title(main=paste("alpha=", as.character
(alphas[j]))
}
```

ここでは、事前分布を仮に決めています、  
実際的にはあいまいな分布の  $\mu$ ,  $\alpha$  を選ぶの  
が妥当。この例では、平均

$$g(\mu) \propto \frac{1}{\mu}, \mu > 0.$$

精度のパラメータ  $\alpha$  を適当にフラットな密度

$$g(\alpha) = \frac{z_0}{(\alpha + z_0)^2}, \alpha > 0.$$

$\alpha$  のメジアンをパラメータ  $z_0$  にとる。ここで  
は  $z_0=0.53$  となる。

## 7.6 Posterior Distribution 事後分布

独立な階層モデルの stage 2 でペアとなる観察  
値の共有する (conjugate) 事前分布を選ぶと  
き、わりと単純な事後分布がある。ハイパー  
パラメータ  $\mu$  と  $\alpha$  の条件下で死亡率  $\lambda_1, \dots, \lambda_{94}$   
は独立の事後分布をとる。  $\lambda_i$  の事後分布は  
 $\text{gamma}(y_i + \alpha, e_i + \alpha/\mu)$ .  $\lambda_i$  の平均は

$$E(\lambda_i | y, \alpha, \mu) = \frac{y_i + \alpha}{e_i + \alpha/\mu} = (1 - B_i) \frac{y_i}{e_i} + B_i \mu,$$

$$B_i = \frac{\alpha}{\alpha + e_i \mu}.$$

真の値  $\lambda_i$  の事後分布平均は収縮項  $B_i$   
(shrinkage fraction), 推定値  $y_i/e_i$  から事前  
平均  $\mu$  に対する縮みで表される。

また、共有する事前分布を使うときは死亡率  
 $\lambda_i$  は joint posterior density を越えて積分さ  
れて  $(\alpha, \mu)$  の marginal posterior density と  
なる。

$$p(\alpha, \mu | \text{data}) = K \frac{1}{\Gamma^{94}(\alpha)} \prod_{j=1}^{94} \left[ \frac{(\alpha/\mu)^\alpha \Gamma(\alpha + y_j)}{(\alpha/\mu + e_j)^{(\alpha + y_j)}} \right] \frac{z_0}{(\alpha + z_0)^2} \frac{1}{\mu},$$

K は定数。

## 7.7 Simulating from the Posterior 事後分布からのシミュレーション

前段で全パラメータの事後分布は

```
g(hyperparameters|data)g(true rates|
hyperparameters, data)
```

と示した。ハイパーパラメータ  $(\alpha, \mu)$ , 真の  
死亡率  $\lambda_1, \dots, \lambda_{94}$  である。 composition  
method によって joint 事後分布からランダム  
抽出する。

- $(\alpha, \mu)$  を周辺事後分布 (marginal  
posterior distribution) からシミュレート
- $\lambda_1, \dots, \lambda_{94}$  をシミュレートした  $\mu$  と  $\alpha$  に基  
づいた分布から求める。

まず、 $\mu$ ,  $\alpha$  ともに正の値であることから実  
数にする。

$$\theta_1 = \log(\alpha), \quad \theta_2 = \log(\mu)$$

これらの周辺事後分布は

$$p(\theta_1, \theta_2 | \text{data}) = K \frac{1}{\Gamma^{94}(\alpha)} \prod_{j=1}^{94} \left[ \frac{(\alpha/\mu)^\alpha \Gamma(\alpha + y_j)}{(\alpha/\mu + e_j)^{(\alpha + y_j)}} \right] \frac{z_0 \alpha}{(\alpha + z_0)^2}.$$

対数  $\theta_1$ ,  $\theta_2$  の事後分布 `poissgamexch` の定  
義

```
poissgamexch=function(theta, datapar)
```

```
{
```

```
y=datapar$[,2]; e=datapar$data[,1]
```

```
z0=datapar$z0
```

```
alpha=exp(theta[,1]); mu=exp(theta[,2])
```

```
beta=alpha/mu
```

```
N=length(y)
```

```
val=0*alpha;
```

```
for (i in 1*n)
```

```
{
```

```
val=val+lgamma(alpha+y[i])-(y[i]+alpha)
```

```
*log(e[i]+beta);
```

```
alpha*log(beta)}
```

```
val=val-N*lgamma(alpha)+log(alpha)-2*log
```

```
(alpha+z0)
```

```
return(val)
```

```
}
```

ただし、この関数には  $2 \times 2$  マトリックス

`theta` とハイパーパラメータを含む `datapar` の

入力が必要。 `theta` はマトリックスなので観測

値の和から指数事後分布を計算する。関数

`lgamma` を使ってガンマ関数  $\log \Gamma(x)$  の指数を

計算する。

`laplace` を使うと事後のモードと分散共分散マ

トリックスを得る。 開始点  $(\theta_1, \theta_2) = (2, -7)$

からNewton-Raphsonアルゴリズムによる反復を行う。

```
datapar=list(data=hearttransplants,
z0=0.53)
start=array(c(2,-7), c(1,2))
fit=laplace(poissgamexch, start, 5,
datapar)
fit
この結果から事後密度の形を推定できる。
いろいろ試行錯誤すると( $\theta_1$ ,  $\theta_2$ )事後密度の
分布が得られる。
par(mfrow=c(1,1))
mycontour(poissgamexch, c(0,8,-7.3,-6.6),
datapar)
title(xlab="log alpha", ylab="log mu")
```

図7.6の事後密度は $\theta_1 = \log \alpha$ 正規分布ではないので適合できない。gibbs関数を使ってMetropolis within Gibbsアルゴリズムで( $\theta_1$ ,  $\theta_2$ )シミュレートする。開始点は( $\theta_1$ ,  $\theta_2$ )=(4,-7), メトロポリスパラメータ $c_1=1$ ,  $c_2=0.15$ 。

```
start<-array(c(4,-7), c(1,2))
fitgibbs <- gibbs(poissgamexch, start,
1000, c(1,.15), datapar)
fitgibbs$accept
結果より, 両方とも許容率約30%
```

図7.7

```
mycontour(poissgamexch, c(0,8,-7.3,
-6.6), datapar)
points(fitgibbs$par[,1], fitgibbs$par[,2])
```

事後分布の精密度パラメータ $\theta_1 = \log(\alpha)$ の密度分布：図7.8

```
plot(density(fitgibbs$par[,1], bw=0.2))
```

事後分布からシミュレートすることによって真の死亡率 $\lambda_1, \dots, \lambda_{94}$ を求めることができる。ハイパーパラメータ $\mu$ と $\alpha$ を与えると真の値は独立な事後分布 $\lambda_i$ は $\text{gamma}(y_i + \alpha, e_i + \alpha/\mu)$ となる。各死亡率についてrgamma関数でサンプリングする。例えば、 $\lambda_1$ の事後分

布からサンプリングする。

```
alpha <- exp(fitgibbs$par[,1])
mu <- exp(fitgibbs$par[,2])
lam1 <- rgamma(1000, y[1]+alpha, e[1]
+alpha/mu)
```

それぞれの $\lambda_i$ について1000のサンプルを得たら、5%と95%点を計算する。この範囲が $\lambda_i$ の90%信頼区間となる。この範囲を図7.2に重ね合わせたのが図7.9。観察された死亡率の幅が大きいのに比べて、真の死亡率の信頼区間はよく似ている。このようにベイズ推定では個々の死亡率が全体の死亡率の近づくように縮んでいる。

```
alpha <- exp(fitgibbs$par[,1])
mu <- exp(fitgibbs$par[,2])
plot(log(e), y/e, pch=as.character(y))
for (i in 1:94){
  lami <- rgamma(1000, y[i]+alpha, e[i]
+alpha/mu)
  probint <- quantile(lami, c(0.05, 0.95))
  lines(log(e[i])*c(1,1), probint)
}
```

## 7.8 Posterior Inferences 事後推定

真の死亡率 $\{\lambda_i\}$ とハイパーパラメータ $\mu$ ,  $\alpha$ がjoint posterior distributionから導かれたら、これを使ってさまざまな推定を行なうことができる。

### 7.8.1 Shrinkage 収縮

$i$ 病院の真の死亡率 $\lambda_i$ の事後平均は

$$E(\lambda_i | \text{data}) \approx (1 - E(B_i | \text{data})) \frac{y_i}{e_i} + E(B_i | \text{data}) \frac{\sum_{j=1}^{94} y_j}{\sum_{j=1}^{94} e_j},$$

$B_i = \alpha / (\alpha + e_i \mu)$ は、 $i$ 病院の観察値 $y_i/e_i$ が全体の推定値 $\sum_{j=1}^{94} y_j / \sum_{j=1}^{94} e_j$ に向かって収縮する程度を表す。次のRコードで、94病院における収縮サイズ $\{B_i\}$ の事後平均を計算する。

```
shrinkage=0*e
for (i in 1:94)shrinkage[i]=mean(alpha/
(alpha+e[i]*mu))
plot (log(e), shrinkage)
```

図7.10は手術数logに対する収縮度をプロットしたもの。ベイズ推定では手術数が少ない病院では90%しか収縮しないが、手術数が増えると50%まで近づけることができる。

## 7.8.2 Comparing Hospitals 病院の比較

個々の病院の真の死亡率から「一番よい病院」を探す場合。まず、いちばん推定値の低い病院を見つける。次のRコードで事後平均死亡率を求める。i病院の事後平均死亡率は

$$E\left(\frac{y_i + \alpha}{e_i + \alpha/\mu}\right)$$

期待値は $(\alpha, \mu)$ での極限事後分布marginal posterior distribution

```
hospital<-1:94
meanrate<-array(0,c(94,1))
for (i in 1:94) meanrate[i]=mean(rgamma
(1000,y[i]+alpha, e[i]+alpha/mu))
hospital[meanrate==min(meanrate)]
```

病院85が死亡率が一番低いこととなる。

次に、i病院とj病院を比べたい場合は $(\lambda_i, \lambda_j)$ の極限分布を求める。そしてi病院の死亡率が低い確率 $P(\lambda_i < \lambda_j)$ を推定値 $(\lambda_i, \lambda_j)$ で $\lambda_i < \lambda_j$ となる割合から求める。次のRコードでこの確率をマトリックスbetterに出力。i病院(行)の要素jがi病院が低い確率。

```
better<-array(0,c(94,94))
for (i in 1:94){
  for (j in (i+1):94){
    if (j <= 94){
      lami=rgamma(1000, y[i]+alpha, e[i]
+alpha/mu)
```

```
lamj=rgamma(1000, y[j]+alpha, e[j]
+alpha/mu)
better[i,j]=mean(lami<lamj)
better[j,i]=1-better[i, j]
}}}
```

病院85と他の病院を比べるにはマトリックスbetterの85列 $P(\lambda_i < \lambda_{85})$ を取り出せばよい。病院1-24の例、病院85が一番死亡率が低いので確率は0に近くなっている。  
> better[1:24, 85]

## 7.9 Posterior Predictive Model Checking 事後推定モデルのチェック

7.3章で死亡率一定モデル $\lambda_1 = \dots = \lambda_{94}$ の適合性を事後予測分布で検定して、このモデルは現状を現していないことを示した。ここでは、同様にexchangeable modelを検定する。

再度、病院94(死亡者17人)を考える。ハイパーパラメータ $\alpha, \mu$ はベクトルalpha, muに代入済み。予測分布 $y^*_{94}$ からまず $\lambda^*_{94}$ の事後分布を求める。

```
lam94<-rgamma(1000,y[94]+alpha, e[94]
+alpha/mu)
```

そして平均 $e_{94} \lambda_{94}$ のポアソン分布から $y^*_{94}$ を求める。

```
ys24<-rpois(1000,e[94]*lam94)
```

図7.11は $y^*_{94}$ のヒストグラムで $y_{94}=17$ の棒を付けたもの。ここでは、観察値はほぼ真ん中に来る。

```
lam94<-rgamma(1000,y[94]+alpha, e[94]
+alpha/mu)
ys94<-rpois(1000,e[94]*lam94)
hist(ys94,breaks=seq(1.5,39.5, by=1))
lines(y[94]*c(1,1), c(0,100), lwd=3)
```

全ての病院について観察値 $y_i$ のあてはまりからexchangeable modelを検定することができる。次のRコードではすべての将来予測 $y^*_i$ が少

なくとも $y_i$ を極限とする確率ベクトル  
pout.exchange.を出力する。

```
pout.exchange <- 0*y
for (i in 1:94) {
  lami=rgamma(1000,y[i]+alpha, e[i]
+alpha/mu)
  ysi=rpois(1000, e[i]*lami)
  pleft=sum(ysi<=y[i])/1000
  pright=sum(ysi>=y[i])/1000
  pout.exchange[i]=min(pleft, pright)
}
```

死亡率一定のときの確率ベクトルがpoutなので、2つのモデルの適合性を比較する（図7.12： $y=x$ の線つき）。

```
plot (pout, pout.exchange, xlab="P
(extreme), equal means", ylab="P
(extreme), exchangeable")
abline(0,1)
```

ここで、exchangeable modelのほうが極限の確率が高いことからフィットがよいと言える。exchangeable modelでは0.1より低い値は2つしかないことから、全般的にも観察値とよく合っている。