第10章　ギブスサンプリング

10.3 プロビットリンクによる2値反応
10.4 平均分割表の推定

茨城大学工学部情報工学科
國井慎也

プロビットリンクによる2値反応の回帰

- 4.4節では、投薬量レベルを関数として、動物の死亡確率の回帰問題を扱った。

  -投薬量レベルの1変数のみ

- これを一般化して、複数の共変量の関数として表すことを考える。

  -プロビット回帰モデル

プロビット回帰のモデル

- ２値の反応：$y_1,\ldots,y_n$　　例（y_i=1なら生存,y_i=0なら死亡）
- 共変量：$x_{i1},\ldots,x_{ik}$　　例（年齢、性別）
- パラメータ：$\beta_1,\ldots,\beta_k$

- プロビット回帰モデルでのy_i=1となる確率p_iは以下で表すことができる

$$p_i = P(y_i = 1) = \Phi(x_{i1}\beta_1 + \cdots + x_{ik}\beta_k)$$

- $\Phi()$ は標準正規分布のcdf(累積分布関数)である

プロビット回帰のモデル(2)

- βの事後分布

  -事前分布に一様事前分布を仮定

$$g(\beta|y) \propto \prod_{i=i}^{n} p_i^{y_i} (1-p_i)^{1-y_i}$$

- 潜在データ

$$Z_i = x_{i1}\beta_1 + \cdot\cdot\cdot + x_{ik}\beta_k + \varepsilon_i$$

  εは、標準正規分布からのランダム標本

- 最終的に、モデルは以下のようになる

$$P(y_i=1) = P(Z_i>0) = \Phi(x_{i1}\beta_1 + \cdot\cdot\cdot + x_{ik}\beta_k)$$

ギブスサンプリングの構成

- ギブスサンプリングアルゴリズムは、潜在データZ_iとパラメータベクトルβの結合事後分布から構成される。

- βの条件付事後分布

$$[\beta | Z, data] \approx N_k((X^{'} X)^{-1} X^{'} Z, (X^{'} X)^{-1})$$

- Z_iの条件付確率（Z_1,..Z_nは独立と仮定）

$$[Z_i | \beta, data] \approx N(x_i \beta, 1) I(Z_i > 0)$$ (y_i=1のとき)

$$[Z_i | \beta, data] \approx N(x_i \beta, 1) I(Z_i < 0)$$ (y_i=0のとき)

Rでの実行(1)

- 使用するデータ

  LearnBayesのdonnerのデータセット

  シエラネバダ山脈を越えようとした移民データ

  大量の餓死者が出た

  →生存確率を求める

- データの内容

  y_i : i番目の生存ステータス(1なら生存,0なら非生存)

  $MALE_i$: 性別(1なら男性,0なら女性)

  AGE_i : 年齢 （15歳以上）

Rでの実行(2)

- プロビットリンクのモデルは以下のようになる

$$P(y_i=1)=\Phi(\beta_0+\beta_1 MALE_i+\beta_2 AGE_i)$$

これに合うようなβを求める

- まずは最尤推定法で求める

- data(donner) #データ読み込み
- > attach(donner)
- > X<-cbind(1,age,male) #デザイン行列Xを作成
- > fit <- glm(survival ~ X - 1, family = binomial(link = probit))
  #最尤推定にはglm関数,βが二項分布に従うのでbinomial,プロビットリンクを指定
- > summary(fit)

Rでの実行(3)

- 最尤推定法出力結果
  - Call:
  - glm(formula = survival ~ X - 1, family = binomial(link = probit))
  - Deviance Residuals:
  -    Min   1Q   Median   3Q   Max
  - -1.7420 -1.0555 -0.2756   0.8861   2.0339
  - Coefficients:
  -       Estimate Std. Error z value Pr(>|z|)
  - X   1.91730   0.76438   2.508   0.0121 *
  - Xage -0.04571   0.02076 -2.202   0.0277 *
  - Xmale -0.95828   0.43983 -2.179   0.0293 *
  - ---
  - Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
  - (Dispersion parameter for binomial family taken to be 1)
  -     Null deviance: 62.383 on 45 degrees of freedom
  - Residual deviance: 51.283 on 42 degrees of freedom
  - AIC: 57.283
  - Number of Fisher Scoring iterations: 5

Rでの実行(4)

- ギブスサンプリングでβを求める

  -bayes.probit関数

  引数:2値反応ベクトル、デザイン行列、サイクル数m

```
> m<-1000

> fit <- bayes.probit(survival, X, m)  #ギブスサンプリングを行う
> apply(fit$beta ,2,mean) #各βの平均を求める
[1]  2.07886032 -0.04954357 -1.04367756

> apply(fit$beta ,2,sd)  # 標準偏差を求める
[1] 0.79242639 0.02078837 0.45506060
```

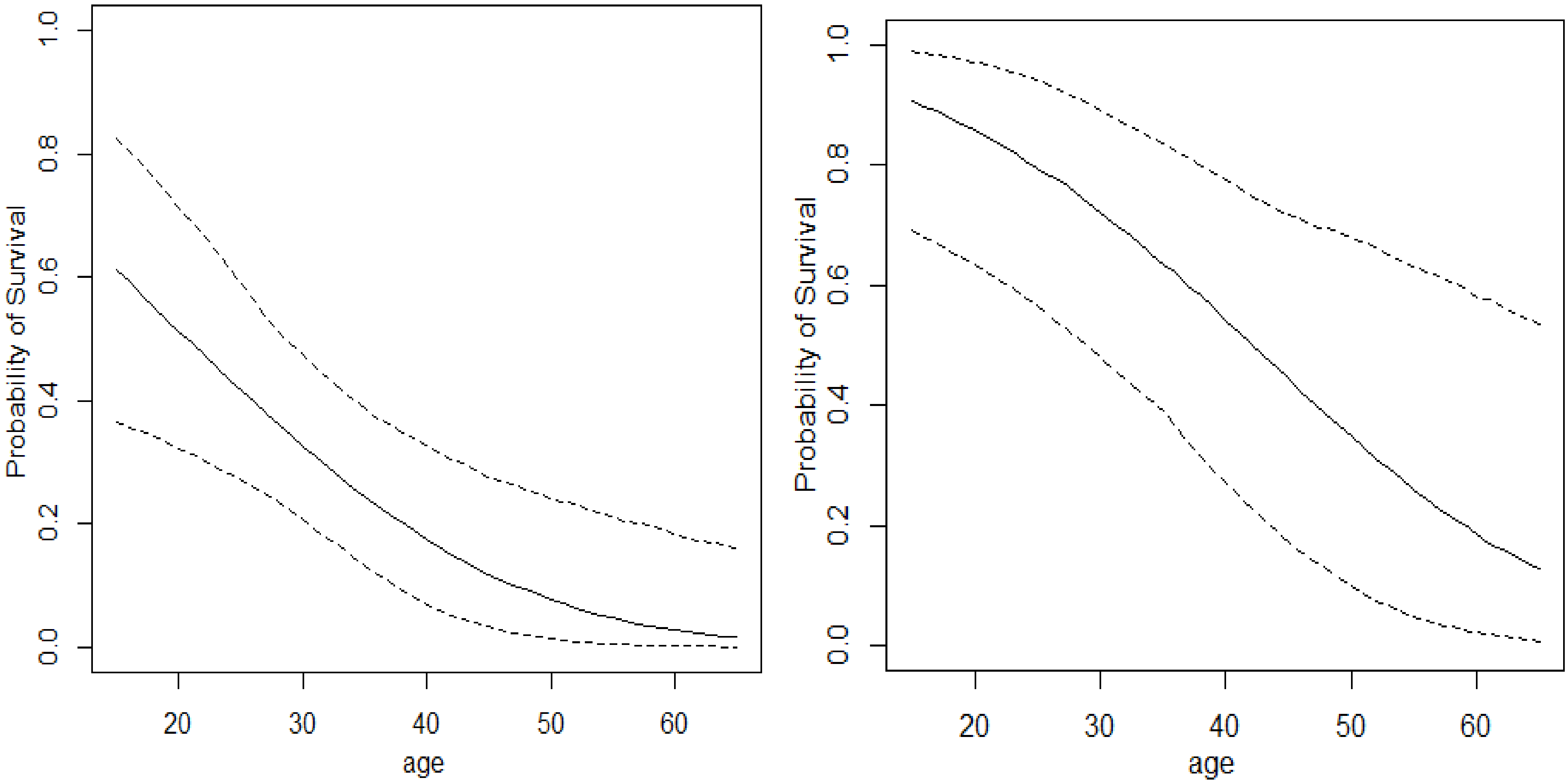
Rでの実行（5）

- 15歳から65歳までの男性のみの生存率を求める

  -bprobit.probs関数：共変量の集合に対する確率の事後分布をシミュレーションした標本から計算する

```
  a=seq(15,65)  #15歳から65歳まで
>  X1=cbind(1,a,1)#共変量からなるベクトルを作成
>  p.male=bprobit.probs(X1,fit$beta)  #男性のみの生存確率を計算する
#以下のコマンドは、グラフを描くため
> plot(a,apply(p.male,2,quantile,.5),type="l",ylim=c(0,1),
+    xlab="age",ylab="Probability of Survival")
> lines(a,apply(p.male,2,quantile,.05),lty=2)
>  lines(a,apply(p.male,2,quantile,.95),lty=2)
```

Rでの実行(6)



左図:男性の生存確率、右図:女性の生存確率

5%,50%,95%分位点をプロット

有情報事前分布

- 先ほどはβに無情報一様事前分布を仮定

  →有情報事前分布を与えることを考える
- 事前分布として、平均ベクトルβ_0、分散共分散行列V_0の多変量正規分布を割り当てる。

- この場合のギブスサンプリングも[β|Z]と[Z|β]からサンプリングされる。βの事前分布は以下のように修正される。

$$[\beta|Z, data] \approx N_k(\beta^1, V_1)$$

$$\beta^1 = (X' X + V_0^{-1})^{-1}(X^{-1} Z + V_0^{-1} \beta_0)$$
$$V_1 = (X' X + V_0^{-1})^{-1}$$

モデル比較

- モデル{性別、年齢}{性別}{年齢}{ゼロモデル}のモデルを比較する。

  比較の基準にベイズファクターを用いる

- ベイズファクター

  ２つのモデルの支持する度合いを計算する

$$BF = \frac{m_1(y)}{m_2(y)} \qquad m(y) = \frac{f(y|\beta)g(\beta)}{g(\beta|y)}$$

f(y|β)：サンプリング密度、g(β):事前密度、f(β|y)事後密度

m(y)の計算

- m(y)の対数をとると

$$\log m(y) = \log f(y|\beta) + \log g(\beta) - \log g(\beta|y)$$

右辺第１項と第２項は既知であるが第３項の計算は難しい

- g(β|y)の計算

βの特定の値をβ*としβ＝β*とおき、潜在データ変数Zを導入すると

$$g(\beta^*|y) = \int g(\beta^*|Z, y)\, g(Z|y)\, dZ$$

m(y)の計算（2）

[β|Z,y]は、正規分布(β_1,V_1)であるので、Z周辺事後密度からg(β*|y)をシミュレーションできる

$$g(\beta^*|y)\approx\frac{1}{m}\sum_{j=1}^{m} g(\beta^*|Z^j, y)=\frac{1}{m}\sum_{j=1}^{m} \Phi(\beta^*;\beta^1, V_1)$$

{Z^j}はシミュレーションによるm個の潜在データの標本集合

Φ(x;μ,V)はxで評価した平均μで分散共分散行列Vの多変量正規密度

- 最終的に、周辺密度の対数の推定値は以下のようになる

$$\log m(y)\approx\log f(y|\beta)+\log g(\beta)-\frac{1}{m}\sum_{j=1}^{m} \Phi(\beta^*;\beta^1, V_1)$$

Rでの実行(7)

- log m(y)を計算する

  bayes.brobit():主観事前分布が使われている場合の事後分布を計算する関数

  引数として

  - y:2値反応ベクトル

  - X:デザイン行列

  - サイクル数

  - list(beta,P)

  - beta:事前平均ベクトル

  - P:事前精度行列

Rでの実行(8)

- data(donner)　　#データ読み込み
- > y=donner$survival　　　#2値反応ベクトルをyに代入
- > X=cbind(1,donner$age,donner$male)　#デザイン行列Xを作成
- >
- > beta0=c(0,0,0); c0=100　#beta0にβの事前平均ベクトルを割り当てる　#c0はβの位置についての確信を与える,数値が大きいほど曖昧な事前知識を反映する
- > P0=t(X)%*%X/c0　# 事前精度行列
- >
- > bayes.probit(y,X,1000,list(beta=beta0,P=P0))$log.marg　#フルモデル
- [1] -31.5153
- >
- > bayes.probit(y,X[,-2],1000,　　　　　　#年齢のみのモデル
- \+　list(beta=beta0[-2],P=P0[-2,-2]))$log.marg
- [1] -32.78796
- >
- > bayes.probit(y,X[,-3],1000,　　　　　　#性別のみのモデル
- \+　list(beta=beta0[-3],P=P0[-3,-3]))$log.marg
- [1] -32.01771

BFの計算

- {年齢、性別}モデルと{年齢}モデルのBFの計算

$$BF = \frac{\exp(-31.5153)}{\exp(-32.78796)} = 3.570337$$

結果的に、{年齢、性別}モデルのほうが支持される

１０．４　はじめに

- この節では、高校での成績から大学での成績を推定をすることを説明する

  HSR：高校での成績ランク

  ACT：高校3年生に課せられる標準化テスト

  GPA：大学での成績

- 使用するデータ

  LearnBayesのiowagpaというデータファイル

  ４０行からなるファイルで、各行にはGPAの標本平均、標本サイズ、高校での百分率での成績ランク、ACTスコアが記録されている

10.4.1　2元分割表の作成

- まず、2元分割表を作成する

  行に、高校での成績ランク（HSR）、列にATCスコア、表の数値に平均GPAを表している

- > library(LearnBayes) #パッケージ読み込み
- > data(iowagpa) ＃データの読み込み

  #以下は、2元表を作成するコマンド
- > rlabels = c("91-99", "81-90", "71-80", "61-70", "51-60", "41-50",
- \+ 　　"31-40", "21-30")
- > clabels = c("16-18", "19-21", "22-24", "25-27", "28-30")
- > gpa = matrix(iowagpa[, 1], nrow = 8, ncol = 5, byrow = T)
- > dimnames(gpa) = list(HSR = rlabels, ACTC = clabels)
- > gpa

	ACTC				
HSR	16-18	19-21	22-24	25-27	28-30
91-99	2.64	3.10	3.01	3.07	3.34
81-90	2.24	2.63	2.74	2.76	2.91
71-80	2.43	2.47	2.64	2.73	2.47
61-70	2.31	2.37	2.32	2.24	2.31
51-60	2.04	2.20	2.01	2.43	2.38
41-50	1.88	1.82	1.84	2.12	2.05
31-40	1.86	2.28	1.67	1.89	1.79
21-30	1.70	1.65	1.51	1.67	2.33

2元分割表の作成(2)

- 次の表では、高校での成績ランクとACTスコアの組み合わせごとに該当する学生数を表している。

```
> samplesizes = matrix(iowagpa[, 2], nrow = 8, ncol = 5, byrow = T)
>  dimnames(samplesizes) = list(HSR = rlabels, ACTC = clabels)
>  samplesizes
```

	ACTC				
HSR	16-18	19-21	22-24	25-27	28-30
91-99	8	15	78	182	166
81-90	20	71	168	178	91
71-80	40	116	180	133	46
61-70	34	93	124	101	19
51-60	41	73	62	58	9
41-50	19	25	36	49	16
31-40	8	9	15	29	9
21-30	4	5	9	11	1

順序制約

- 成績における順序制約

  GPAは、次の２つの順序制約を満たすと考えるのが妥当

  u_{ij}:HRSのレベルがi、ACTスコアのレベルがjの学生の母集団のGPAの平均

$$u_{i1} \leq u_{i2} \leq \cdot\cdot\cdot \leq u_{i5}$$

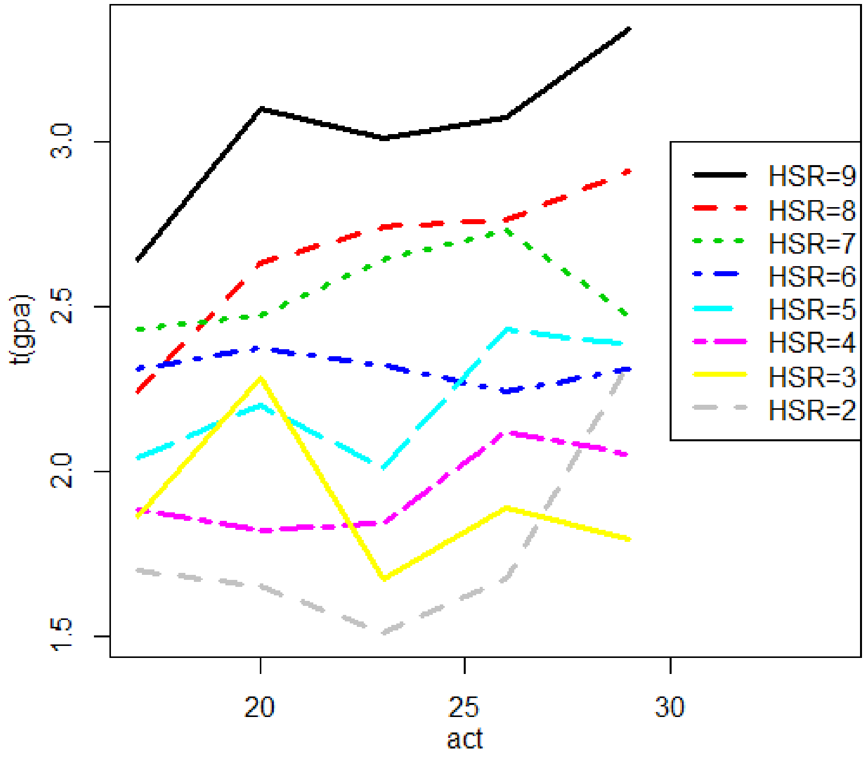
  これは、ACTスコアが高くなればなるほどGPAも高くなるという確信を表す

$$u_{1\mathrm{j}} \leq u_{2\mathrm{j}} \leq \cdot\cdot\cdot \leq u_{9\mathrm{j}}$$

  これは、HSRスコアが高くなればなるほどGPAも高くなるという確信を表す

```
act = seq(17, 29, by = 3)  #x軸の範囲を指定
matplot(act, t(gpa), type = "l", lwd = 3,  #以下、グラフを作成するコマンド
 xlim = c(17, 34), col=1:8, lty=1:8)
legend(30, 3, lty = 1:8, lwd = 3, legend = c("HSR=9", "HSR=8",
    "HSR=7", "HSR=6", "HSR=5", "HSR=4", "HSR=3", "HSR=2"), col=1:8)
```

- 高校での成績ランク(HSR)とACTスコアのレベルごとのGPA平均
- 順序制約を完全に満たしていないので、順序制約に適合するよう母集団平均を平滑化した推定値を求めることが望ましい



10.4.2 順序制約に従うモデル

- 母集団平均は、順序制約に従うが、平均の位置については知識がないとすると、

  事前分布は、
  $$g(\mu) \propto c(\mu \in A)$$

  Aは、順序制約にしたがうμの値の空間を意味する

- μの尤度関数

  y_{ij}:GPAの標本平均,n_{ij}：標本サイズ

  観測値y_{ij}は、それぞれ独立で、平均μ_{ij},分散σ^2/n_{ij}の正規分布に従うと仮定。σは既知とすると

$$L(\mu) = \prod_{i=1}^{8} \prod_{j=1}^{5} \exp\left\{ -\frac{n_{ij}}{2\sigma^2} (y_{ij} - \mu_{ij})^2 \right\}$$

順序制約に従うモデル（2）

- 事後密度は以下で与えられる

$$g\left(\mu|y\right) \propto L\left(\mu\right)$$

- 次に、ギブスサンプリングを行う

  ordergibbs関数：順序制約に従うギブスサンプリングを行う

  引数

  標本平均と標本サイズを含む行列

  サイクル数

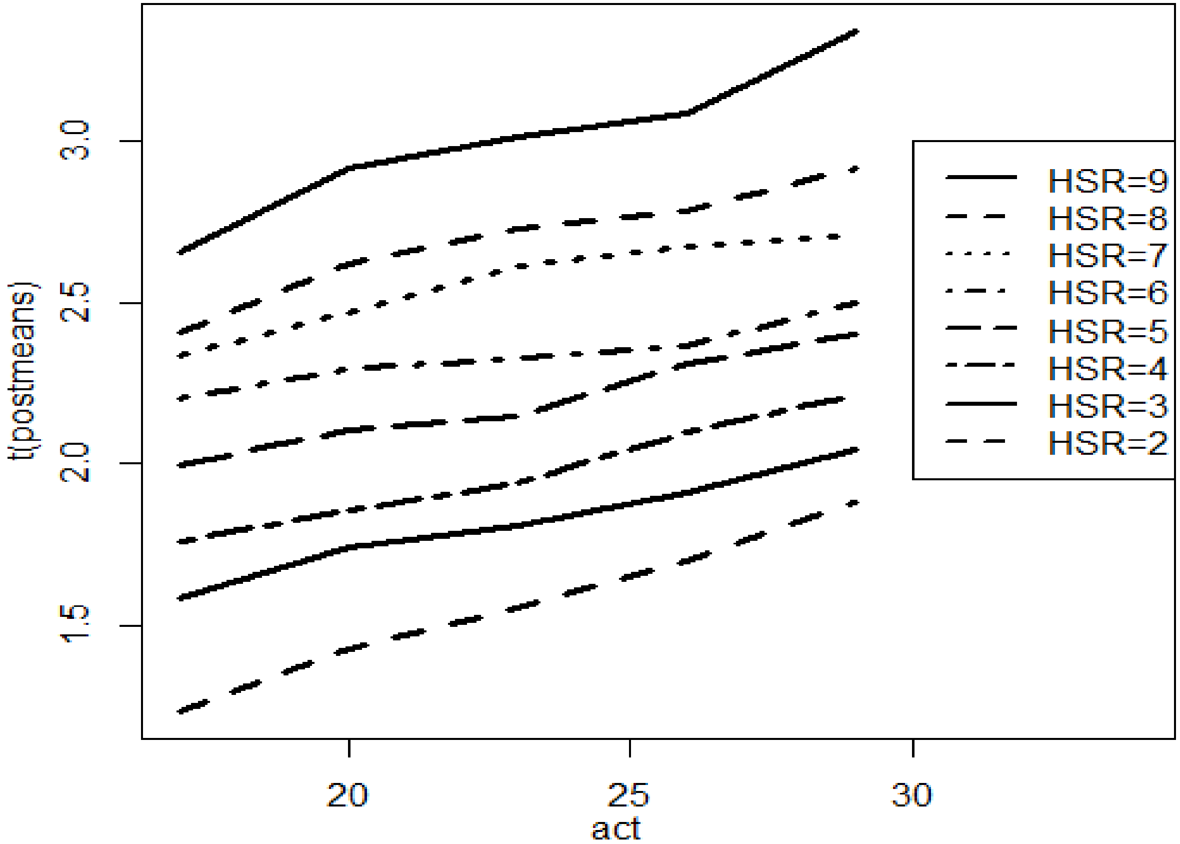
GPAの事後平均を求める

- ordergibbs関数を使って、サンプリングを行い、GPAの事後平均を求める。サイクル数は5000回。
  - > MU = ordergibbs(iowagpa, 5000) #順序制約に従ったギブスサンプリングを行う
  - > postmeans = apply(MU, 2, mean) #求めた標本の平均を求める
  - > postmeans = matrix(postmeans, nrow = 8, ncol = 5) ##以下、表を作成するコマンド
  - > postmeans=postmeans[seq(8,1,-1),]
  - > dimnames(postmeans)=list(HSR=rlabels,ACTC=clabels)
  - > round(postmeans,2)

	ACTC				
HSR	16-18	19-21	22-24	25-27	28-30
91-99	2.65	2.92	3.01	3.09	3.34
81-90	2.41	2.62	2.73	2.78	2.92
71-80	2.33	2.47	2.62	2.67	2.71
61-70	2.20	2.29	2.33	2.37	2.50
51-60	2.00	2.11	2.15	2.31	2.40
41-50	1.76	1.86	1.94	2.10	2.21
31-40	1.59	1.74	1.81	1.91	2.05
21-30	1.24	1.43	1.56	1.70	1.88

GPAの事後平均を求める（2）

```
matplot(act, t(postmeans), type = "l", lty=1:8, lwd = 3, col = 1, xlim = c(17, 34))
legend(30, 3, lty = 1:8, lwd = 2, legend = c("HSR=9", "HSR=8",
    "HSR=7", "HSR=6", "HSR=5", "HSR=4", "HSR=3", "HSR=2"))
```



- 順序制約空間に無情報事前分布を使ったGPAの事後平均プロット

  順序制約に従っているのがわかる

順序制約が与える影響

- 順序制約という事前の確信が推論に与える影響を調べたい
  →セル平均事後標準偏差を計算し、古典的な標準偏差と比べること
- まず、事後標準偏差を求める

```
• postsds = apply(MU, 2, sd)  #平均の事後標準偏差を求める
• >  postsds = matrix(postsds, nrow = 8, ncol = 5) ##以下、表を作成するコマンド
• >  postsds=postsds[seq(8,1,-1),]
• >  dimnames(postsds)=list(HSR=rlabels,ACTC=clabels)
• >  round(postsds,3)
```

```
       ACTC
HSR     16-18 19-21 22-24 25-27 28-30
  91-99 0.140 0.085 0.054 0.044 0.051
  81-90 0.078 0.059 0.038 0.038 0.063
  71-80 0.065 0.052 0.039 0.039 0.046
  61-70 0.066 0.040 0.037 0.039 0.081
  51-60 0.073 0.053 0.055 0.049 0.073
  41-50 0.082 0.067 0.067 0.071 0.086
  31-40 0.116 0.079 0.073 0.074 0.096
  21-30 0.184 0.139 0.120 0.112 0.131
```

順序制約が与える影響（2）

- 古典的な観測標本平均y_{ij}の標準誤差はSE(yij)=σ/√n_{ij} で与えられる。
- 次の表は、すべてのセルについての比{SD(μij|y)/SE(uij)} を計算している。

```
> s=.65  #σ=0.65と仮定
>  se=s/sqrt(samplesizes) #SEを計算
>  round(postsds/se,2)  #SD/SEを計算し、表を作成する
```

	ACTC				
HSR	16-18	19-21	22-24	25-27	28-30
91-99	0.61	0.51	0.74	0.90	1.00
81-90	0.54	0.76	0.75	0.77	0.92
71-80	0.63	0.86	0.80	0.69	0.48
61-70	0.59	0.59	0.63	0.60	0.54
51-60	0.72	0.69	0.66	0.57	0.34
41-50	0.55	0.52	0.62	0.77	0.53
31-40	0.50	0.36	0.43	0.61	0.44
21-30	0.57	0.48	0.55	0.57	0.20

- ほとんどのセルの比が0.5から0.7にあるので、順序制約のある事前分布を利用したことで、有意な事前情報が加えられたとを示している。

10.4.3 階層的回帰モデル

- GPAと二つの説明変数をの関係を反映するために階層的回帰事前分布を構成する
- μの分布

  平均が独立で、μ_{ij}が以下の回帰構造で与えられる位置を持ち分散が σ_{π}^{2} の正規分布に従うと仮定する

  $$\beta_0 + \beta_1 ACT_i + \beta_2 HSR_j$$

- 超パラメータβの分布

  $\beta = (\beta_0, \beta_1, \beta_2)$ は $N_3(\overline{\beta}, \Sigma_\beta)$ に従うと仮定

  $\overline{\beta}$は、パラメータの平均、Σ_βは分散共分散行列

階層的回帰モデル（２）

- ${\sigma_\pi}^2$ の分布

  ${\sigma_\pi}^2$ は S_{Xv}^{-2} に従うと仮定する

  S_{Xv}^{-2} ：自由度vの逆カイ2乗分布

- パラメータ(μ,β, ${\sigma_\pi}^2$)の事後密度は以下のようになる

$$g(\mu, \beta, {\sigma_\pi}^2 | y) \propto \prod_{i=1}^{8} \prod_{j=1}^{5} \exp\left\{ -\frac{n_{ij}}{2\sigma^2} (y_{ij} - \mu_{ij})^2 \right\} \times \prod_{i=1}^{8} \prod_{j=1}^{5} \frac{1}{\sigma_\pi} \exp\left\{ -\frac{n_{ij}}{2{\sigma_\pi}^2} (\mu_{ij} - x_{ij}{}' \beta)^2 \right\}$$

$$\times \exp\left\{ -\frac{1}{2} (\beta - \overline{\beta}) \sum^{-1} (\beta - \overline{\beta}) \right\} ({\sigma_\pi}^2)^{-v/2-1} \exp\left\{ -\frac{S}{2{\sigma_\pi}^2} \right\}$$

パラメータの推定

- このモデルのパラメータ(μ,β, σ_π^2) の結合事後密度からシミュレーションするには、hiergibbs関数を用いる

  hiergibbs関数:階層的回帰モデルでギブスサンプリングを行う

  引数：dataファイル、サイクル数

```
FIT=hiergibbs(iowagpa,5000)
```

βの推定

- Bの値の分位点を求める

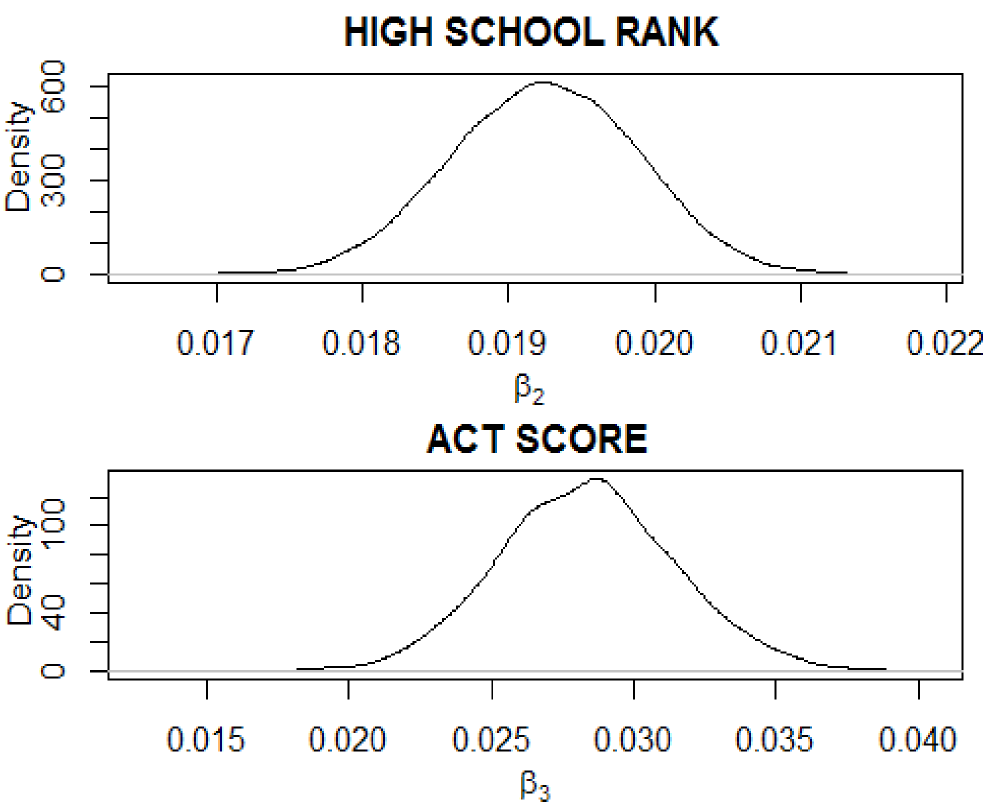
  分位点を求めるにはquantile関数を用いる

```
> quantile(FIT$beta[,2],c(.025,.25,.5,.75,.975)) #β1の分位点を求める
      2.5%        25%        50%        75%      97.5%
0.01799428 0.01880991 0.01925748 0.01968191 0.02046982
>
> quantile(FIT$beta[,3],c(.025,.25,.5,.75,.975)) #β2の分位点を求める
      2.5%        25%        50%        75%      97.5%
0.02238404 0.02624022 0.02834135 0.03034493 0.03455294
>
```

- β_1の95%区間推定値は（0.017,0.020）
- β_2の95%区間推定値は（0.022,0.034）

βの推定（2）

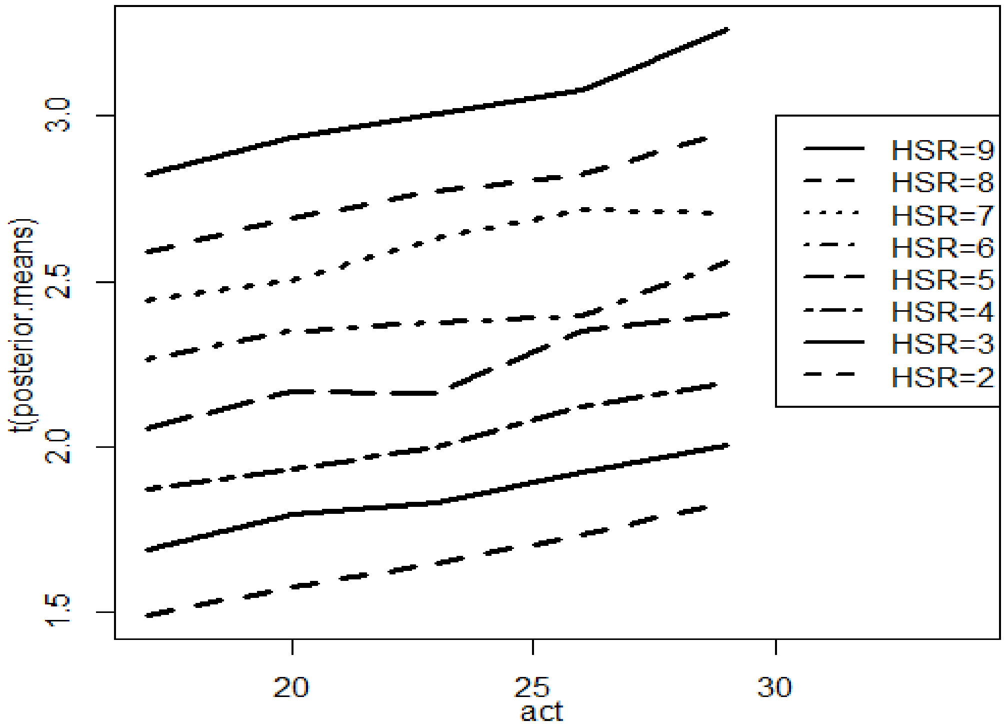
```
 par(mfrow=c(2,1))  ##以下、グラフを描くためのコマンド
>  plot(density(FIT$beta[,2]),xlab=expression(beta[2]),
+   main="HIGH SCHOOL RANK")
>  plot(density(FIT$beta[,3]),xlab=expression(beta[3]),
+   main="ACT SCORE")
```



- 階層的回帰モデルにおける回帰係数β_2とβ_3のシミュレーションした標本の密度推定値

GPAの事後平均を求める

```
• posterior.means = apply(FIT$mu, 2, mean)  #GPAの事後平均を求める
• >  posterior.means = matrix(posterior.means, nrow = 8, ncol = 5,  #以下、グラフを描くためのコマンド
• +   byrow = T)
• >  par(mfrow=c(1,1))
• >  matplot(act, t(posterior.means), type = "l", lwd = 3, lty=1:8, col=1,
• +    xlim = c(17, 34))
• >  legend(30, 3, lty = 1:8, lwd = 2, legend = c("HSR=9", "HSR=8",
• +      "HSR=7", "HSR=6", "HSR=5", "HSR=4", "HSR=3", "HSR=2"))
```



- 階層的事前モデルを使ったGPAの事後平均のプロット図

10.4.4 学生の成績を予測する

- 大学は、将来入学してくる学生の成績を予測したい

  GPAの2.5以上が成功したとするならば以下の予測事後密度を計算する

$$P\left(z_{ij}^{*} \geq 2.5 \middle| y\right)=\int P\left(z_{ij}^{*} \geq 2.5 \middle| \mu, y\right) g\left(\mu \middle| y\right) d\mu$$

  z_{ij}^*:ACTスコアがiで、高校の成績ランクがjの学生が将来入学する大学のGPAとする

  $g(\mu|y)$:セル平均のベクトルμの事後分布

- z_{ij}の分布を正規分布$N(\mu_{ij},\sigma)$に従うと仮定すると

$$P\left(z_{ij}^{*} \geq 2.5 \middle| y\right)=\int \left[1-\Phi\left(\frac{2.5-\mu_{ij}}{\sigma}\right)\right] g\left(\mu \middle| y\right) d\mu$$

学生の成績を予測する(2)

- 仮に $\{u_{ij}{}^{t}, t=1,\ldots,m\}$ がu$_{ij}$の周辺事後分布から標本だとすると

$$P\left(z_{ij}^{*} \geq 2.5|y\right) \approx \frac{1}{m}\sum_{t=1}^{m}\left[1-\Phi\left(\frac{2.5-\mu_{ij}{}^{t}}{\sigma}\right)\right]$$

- 予測確率を計算する

hiergibbs関数の出力はFITであるので、FIT$muが事後分布からシミュレーションされたセル平均の行列となる

```
 p=1-pnorm((2.5-FIT$mu)/.65) #上の式の[ ]内を計算する
>  prob.success=apply(p,2,mean) #確率pの平均を計算する
> #以下、表を表示させるコマンド
>  prob.success=matrix(prob.success,nrow=8,ncol=5,byrow=T)
>  dimnames(prob.success)=list(HSR=rlabels,ACTC=clabels)
>  round(prob.success,3)
```

学生の成績を予測する（3）

ACTC HSR	16-18	19-21	22-24	25-27	28-30
91-99	0.690	0.748	0.781	0.813	0.879
81-90	0.554	0.617	0.662	0.690	0.757
71-80	0.466	0.503	0.579	0.630	0.625
61-70	0.360	0.411	0.425	0.438	0.537
51-60	0.249	0.304	0.304	0.410	0.441
41-50	0.168	0.193	0.221	0.281	0.319
31-40	0.107	0.140	0.153	0.189	0.224
21-30	0.062	0.078	0.096	0.121	0.153

- 将来のGPAが2.5を超える予測確率