

Rで学ぶベイズ統計学入門

7.6 事後分布

7.7 事後分布からのシミュレーション

7.8 事後推論

7.9 ベイズ感度分析

7.10 事後予測モデルの検証

茨城大学工学部情報工学科
國井慎也

事後分布(1)

- 階層モデルが独立構造をしていること、第二段階での共役事前密度の選択により、事後分析は比較的単純になる。

- λ_i の事後平均

超パラメータ μ 、 α を条件

死亡率 $\lambda_1, \dots, \lambda_{94}$ は互いに独立である事後分布に従う

λ_i の事後分布は、 γ 分布 $\text{gamma}(y_i + \alpha, e_i + \alpha/\mu)$ に従う

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

- 真の死亡率 λ_i の事後平均は、縮約推定量と考えられる

B_i は、事後平均を推定量 y_i/e_i から事前平均 μ に近づける縮約係数

事後分布(2)

- (α, μ) の周辺事後密度

共役モデル構造が使われているので、死亡率 λ_i は結合事後密度から積分消去できる

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

Kは比例定数

事後分布からシミュレーション(1)

- 前節では、全パラメータの事後密度は
 $g(\text{超パラメータ}|\text{data})g(\text{真の率}|\text{超パラメータ},\text{data})$
超パラメータ: (α, μ)
真の率: $(\lambda_1, \dots, \lambda_{94})$
- 次のように結合事後密度からランダムな標本をシミュレーションする
 - ・周辺事後分布から (μ, α) をシミュレーションする
 - ・求めた (μ, α) を条件として、 $\lambda_1, \dots, \lambda_{94}$ をそれらの分布からシミュレーションする

事後分布からシミュレーション(2)

まず, μ , α ともに正の値であることから実数値のパラメータに変換する。

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

変換された周辺事後分布は

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

事後分布からシミュレーション(3)

- θ_1 と θ_2 の対数事後分布を求める(poissgamexch関数)
- laplace関数を用いて、事後モードを求める。

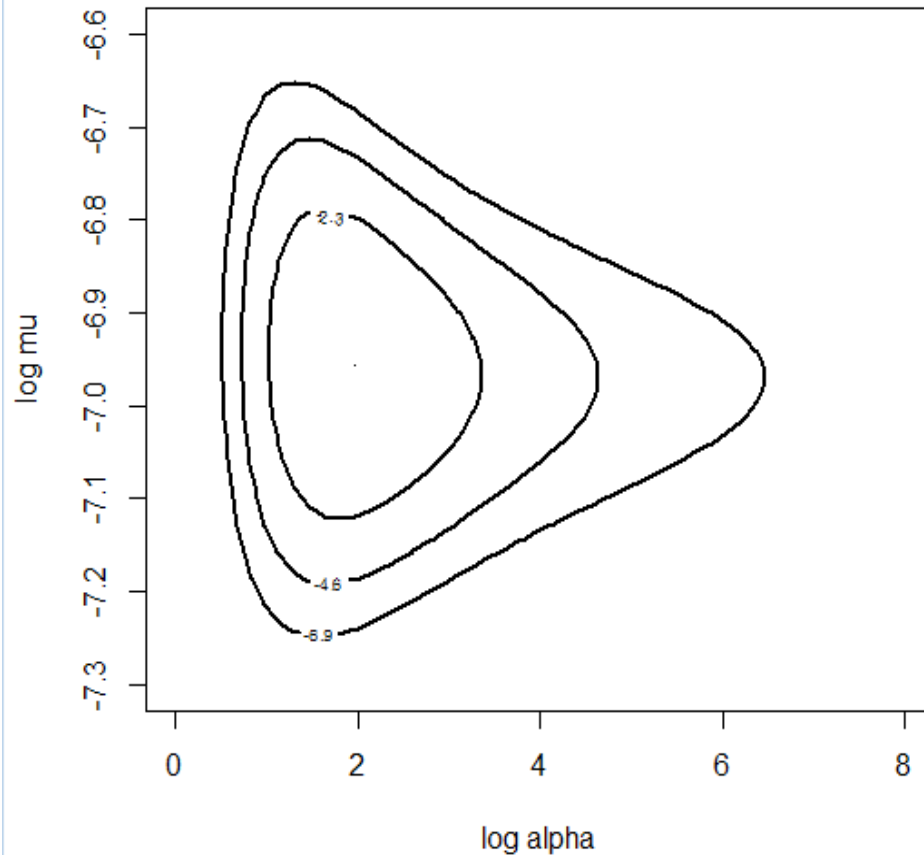
初期値 $(\theta_1, \theta_2) = (-2, 7)$

```
• $mode
• [1] 1.883954 -6.955446
•
• $var
•      [,1]      [,2]
• [1,] 0.233694921 -0.003086655
• [2,] -0.003086655 0.005866020
•
• $int
• [1] -2208.503
•
• $converge
• [1] TRUE
```

モードが $(\theta_1, \theta_2) = (1.88, -6.95)$ となる

事後分布からシミュレーション(4)

- mycontour関数で等高線を描く



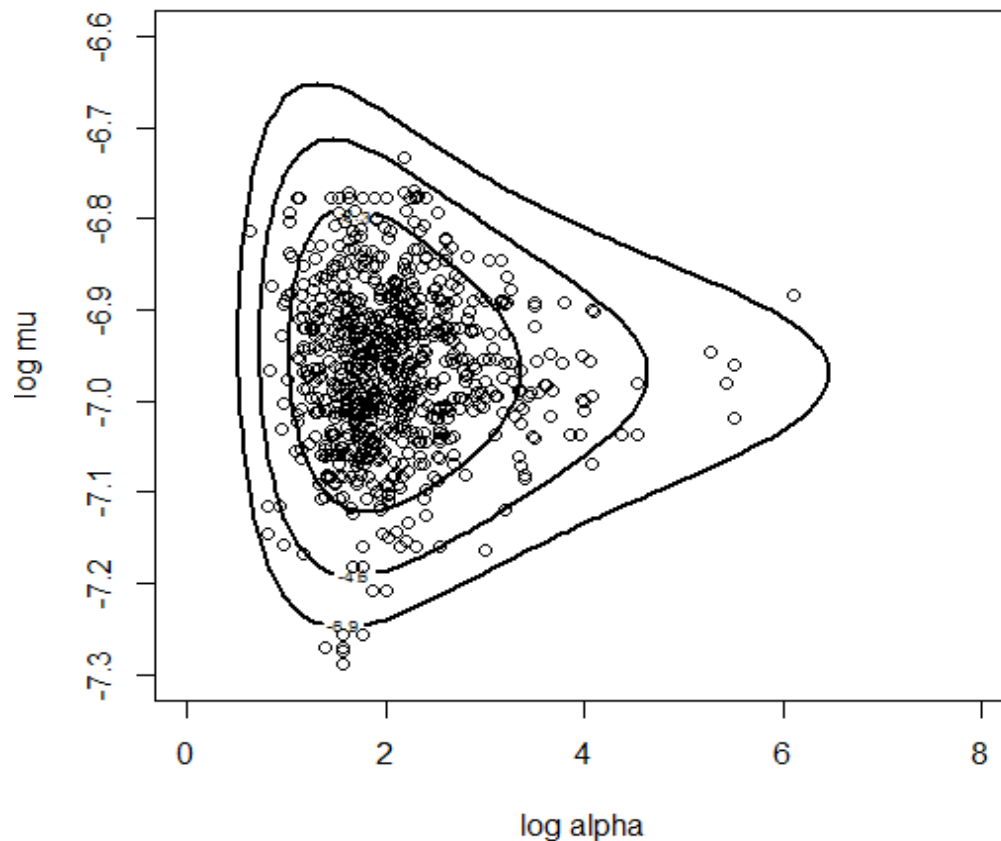
- (θ_1, θ_2) の事後密度は正規形ではない。
よって、正規近似するのは、不適。
- gibbs関数のギブス内メトロポリスアルゴリズムを使って、標本を抽出する。

事後分布からシミュレーション(5)

初期値 $(\theta_1, \theta_2) = (4, -7)$

尺度パラメータ $(c_1, c_2) = (1, 0.15)$

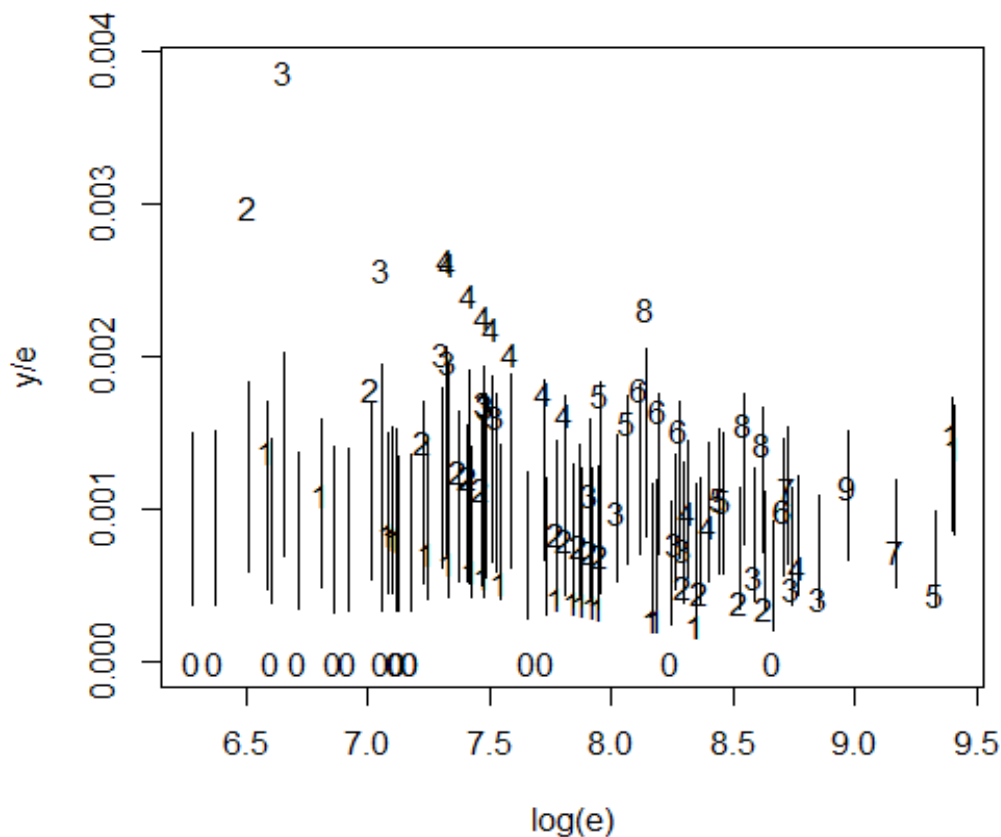
```
>fitgibbs<-gibbs(poissgamexch,start,1000,c(1,.15),datapar)
```



事後分布からシミュレーション(6)

- (α, μ) からの真の死亡率 $\lambda_1, \dots, \lambda_{94}$ を検討する。

真の死亡率は、ガンマ分布 $\text{gamma}(y_i + \alpha, e_i + \alpha/\mu)$ に従う



- 真の死亡率の90%区間を線で表している。
- 真の死亡率の区間の位置は、それぞれ似ていることがわかる。

縮約(1)

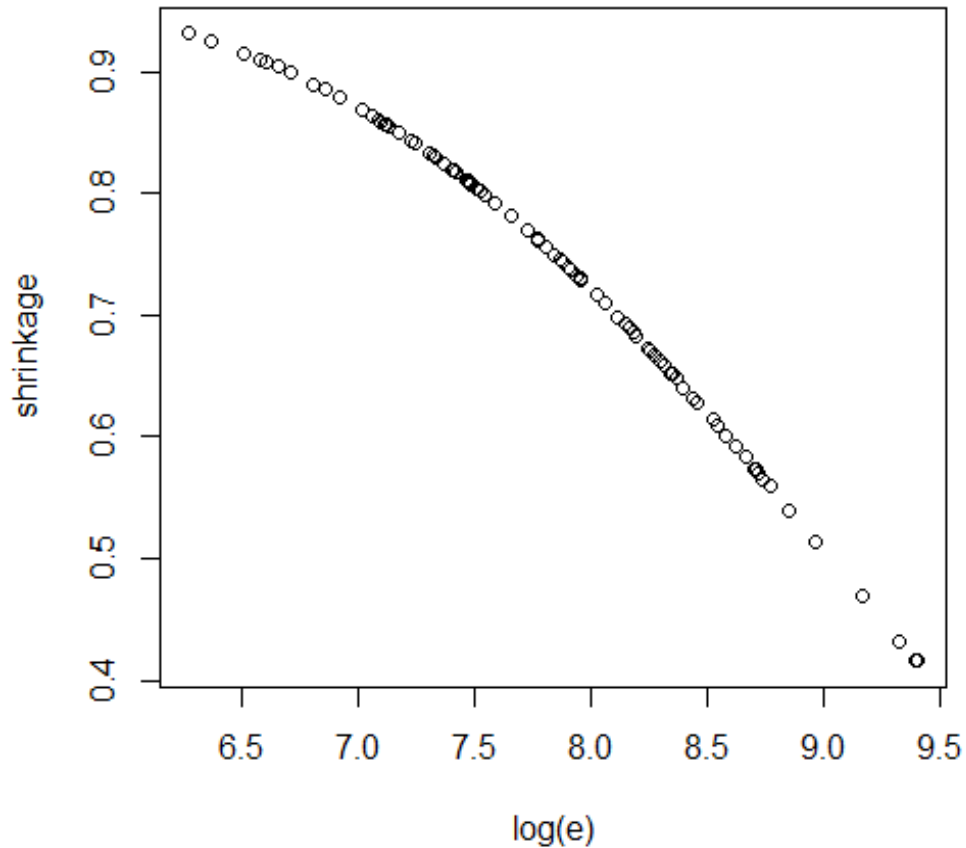
- 病院*i*の真の死亡率 λ の*i*事後平均は、以下ので近似可能である。

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

- $B_i = \alpha / (\alpha + e_i \mu)$ は、プールされた推定値 $\frac{\sum_{j=1}^{94} y_j}{\sum_{j=1}^{94} e_j}$ への病院*i*の観測死亡率 y_i/e_i の縮約サイズである。
- B_i の事後平均を計算する。

縮約(2)

- B_i の縮約サイズをプロットした図



- 暴露数が少ない病院は、個々の死亡率を結合死亡率に90%縮約する。
- 暴露数が多くなるほど、縮約される率は下がる。

病院間の比較(1)

- 真の死亡率を病院間で比較する。
特に「最良の病院」を見つけ、それを他の病院と比較する。
- まず、最良の病院を見つける

病院*i*の真の死亡率の事後平均 $E\left(\frac{y_i + \alpha}{e_i + \frac{\alpha}{\mu}}\right)$ で与えられる

```
• s> mrate <-function(i)mean(rgamma(1000,y[i]+alpha,e[i]  
• + +alpha/mu))  
• > hospital<-1:94  
• > meanrate<-sapply(hospital,mrate)  
• > hospital[meanrate==min(meanrate)]  
• [1] 85
```

- 最良な病院は85となる

病院間の比較(2)

- 2つの病院*i,j*間で死亡率を比較する。
 λ_i, λ_j を生成し、確率 $P(\lambda_i < \lambda_j)$ となる確率を求める。
- ここでは、病院85と他の病院の死亡率を比較する。

```
• sim.lamda<-function(i) rgamma(1000,y[i]+alpha,e[i]+alpha/mu)
• > LAM<-sapply(1:94,sim.lamda)
• > compare.rates<-function(x)
• + {
• + nc <-NCOL(x)
• + ij <-as.matrix(expand.grid(1:nc,1:nc))
• + m<-as.matrix(x[,ij[,1]]>x[,ij[,2]])
• + matrix(colMeans(m),nc,nc,byrow=TRUE)
• + }
• > better<-compare.rates(LAM)
• > better[1:24,85]
• [1] 0.188 0.200 0.084 0.139 0.145 0.234 0.219 0.182 0.054 0.232 0.205 0.183
• [13] 0.212 0.106 0.068 0.207 0.249 0.096 0.271 0.141 0.145 0.141 0.061 0.084
```

- 求めた確率が0に近いので病院85が最良な病院だとわかる

ベイズ感度分析(1)

- ベイズ分析ではモデルの仮定に変更を加えた場合の推論の感度を評価することは重要である。
- ここでは、事前密度 $g(\alpha) = z_0/(\alpha+z_0)^2$ の z_0 の値が変更になるとき α の事後分布の感度を評価する。

$\theta_1 = \log \alpha$ として、 θ_1 の事前分布は

$$g(\theta_1 | z_0) = \frac{z_0 \exp(\theta_1)}{(z_0 + \exp(\theta_1))^2}$$

- ここで、 $z_0 = 0.53$ ではなく $z_0 = 5$ とするときの $\log \alpha$ の事後分布を評価する

ベイズ感度分析(2)

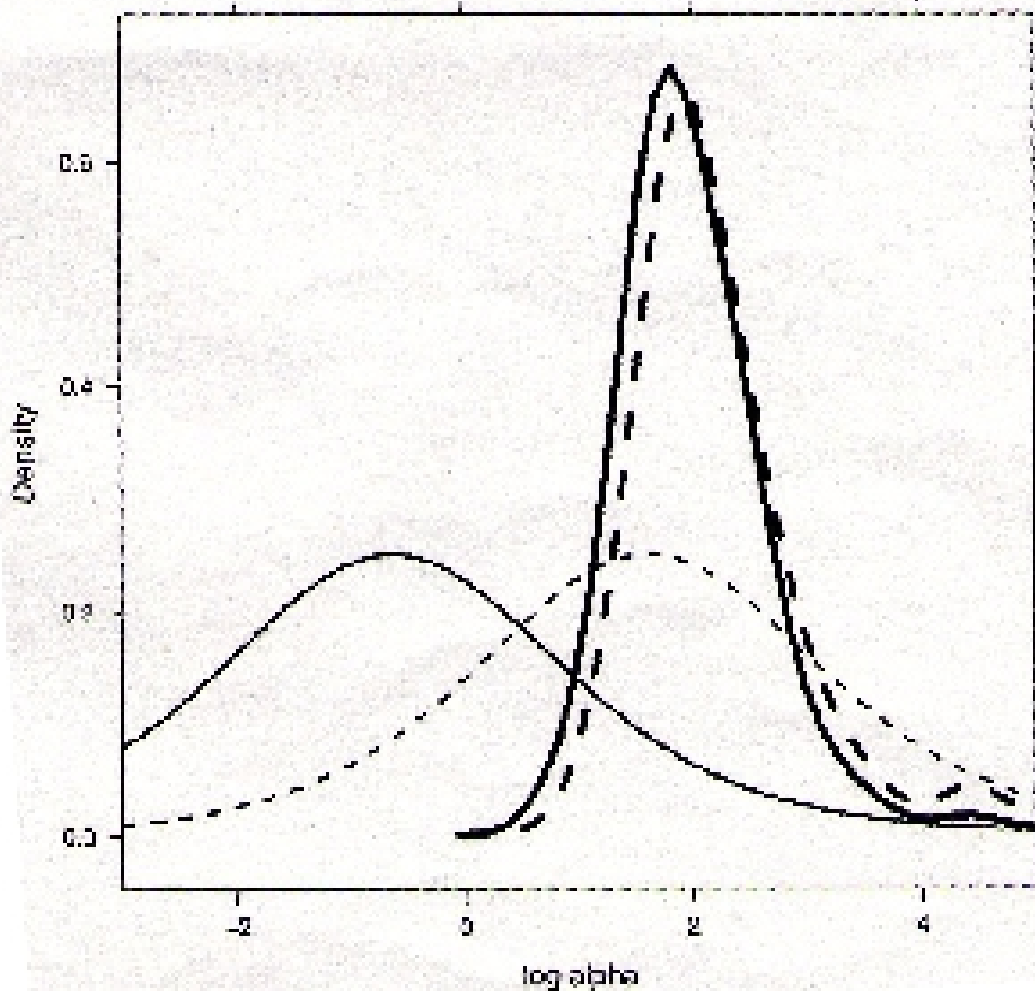
- ある事後分布から θ_1 をシミュレーションした標本を新しい分布に変換するためにSIRアルゴリズムを使う。

重みは、

$$w(\theta_1) = \frac{g(\theta_1 | z_0 = 5)}{g(\theta_1 | z_0 = 0.53)}$$

ベイズ感度分析(3)

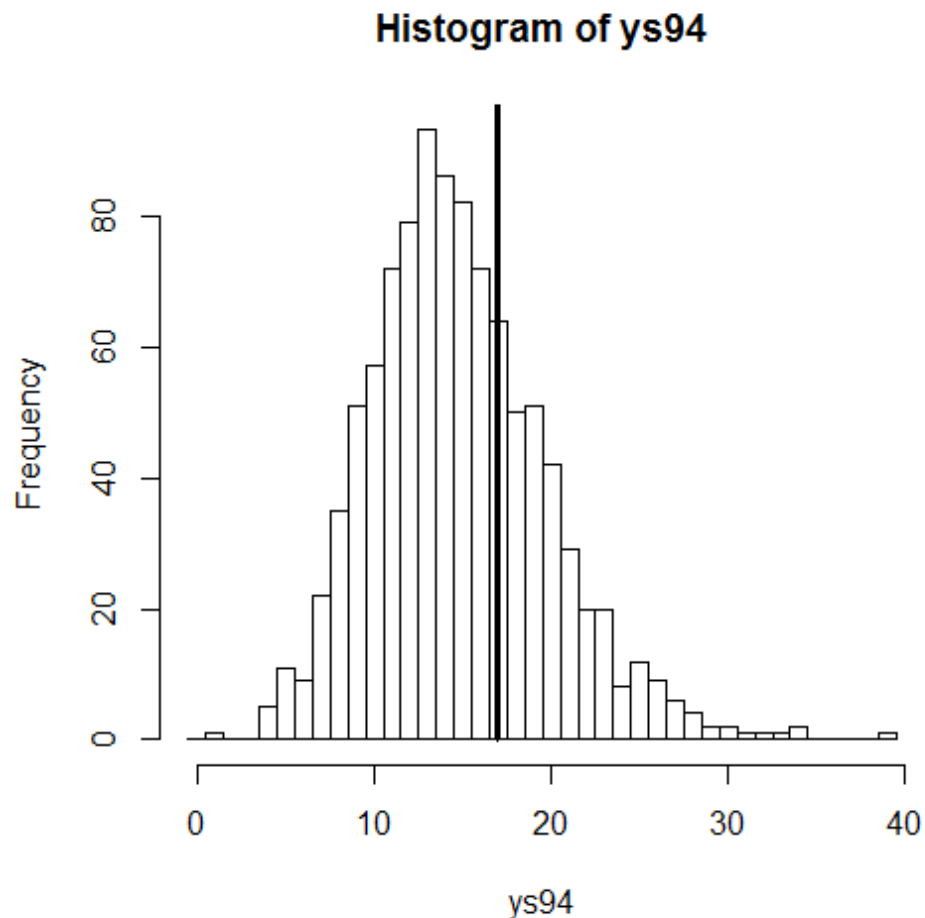
Original Prior and Posterior (solid),
New Prior and Posterior (dashed)



- 細い実線と破線は、それぞれの元の事前密度と新しい事前密度
- 太い実線と破線は、それぞれの事後密度を表している。
- Z_0 の選択は、モデルの事後密度にさほど影響を与えない。

事後予測モデルの検証(1)

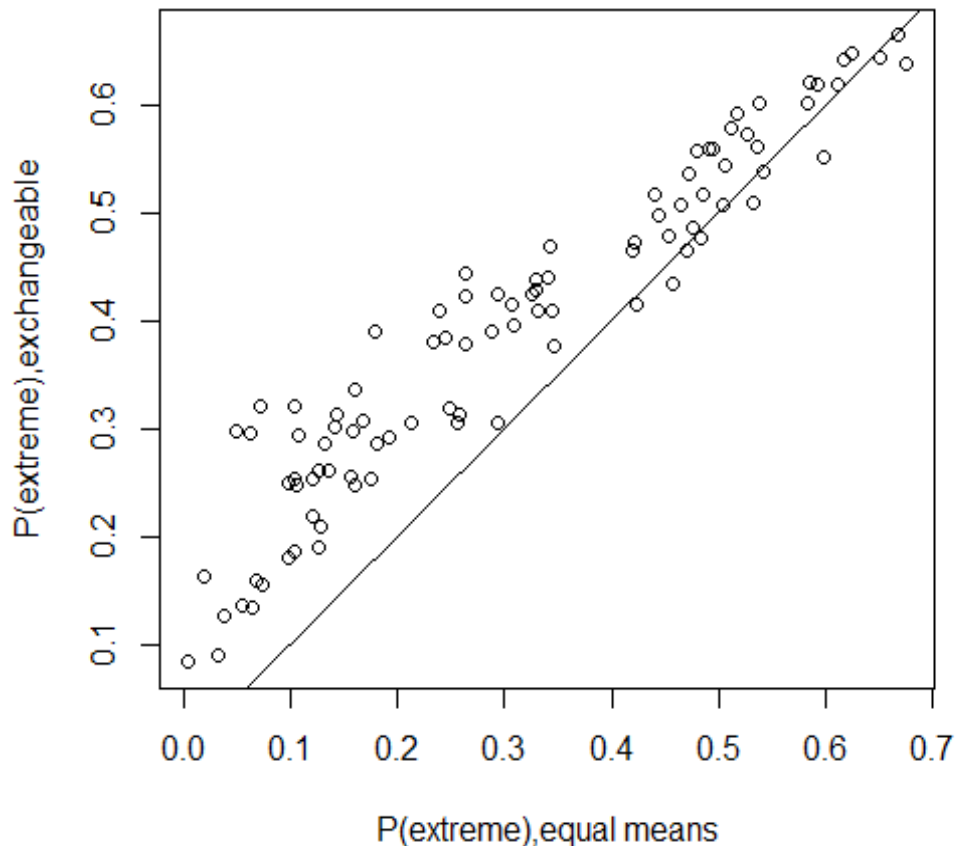
- 7.3節の事後予測モデルを使って、交換可能性モデルを検証する。



- y_{94}^* の事後予測分布に、 y_{94} の実測値を重ねたもの。
- 病院94で観測された死亡数は、予測分布の中央に位置してる。よって、観測データは、モデルに合致している。

事後予測モデルの検証(2)

- 等平均モデルと交換可能モデルそれぞれの「少なくとも同程度に極端な裾に位置する確率」の事後予測確率の散布図



- 交換可能モデルのほうが裾の外れに位置する確率が大きい
- 観測値が矛盾しないのは、交換可能モデルの方である。