Rで学ぶベイズ統計学入門

第7章
階層モデリング

7.1 はじめに　　7.2 3つの事例
7.3 個々の推定と組み合わせ推定
7.4 死亡率は等しいか?
7.5 交換可能性を事前の確信とするモデリング

新納 浩幸

階層モデリング

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

g_2 により λ を生成

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

λ の下で

g_1 により Θ を生成

$$y \sim f(y|\theta)$$

Θ の下で

f により y を生成

見えているのはこのデータのみ

例題(病院死亡率の同時推定)

94 の病院における各病院の死亡率 $\lambda_i \quad (i = 1 \sim 94)$
を推定したい

手術Aを行った結果の死亡率・・・?

使えるデータ

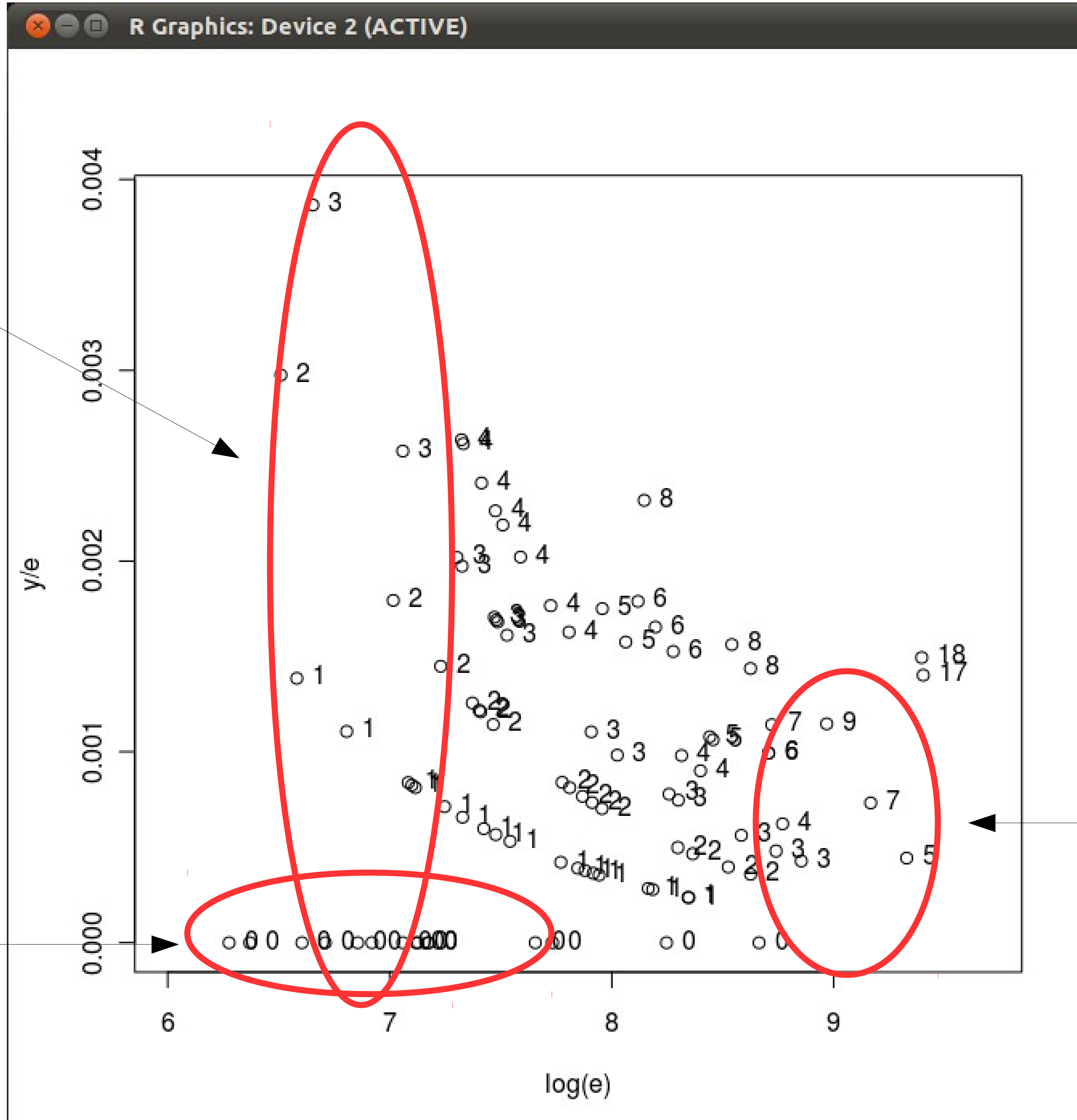
y_i 病院 i での(手術A後の)死亡数

e_i 病院 i での手術Aの実施数

$$\lambda_i = \frac{y_i}{e_i}$$

通常はこれでOK、だけど
データ数が少ないと怪しい

分散大


R Graphics: Device 2 (ACTIVE)
y/e
0.000
0.001
0.002
0.003
0.004
log(e)
6
7
8
9

分散小

y/e = 0

死亡率は等しい?

$$\lambda_1 = \lambda_2 = \cdots = \lambda_{94} = \lambda$$

$$\Rightarrow \quad \lambda = \frac{\sum_{j=1}^{94} y_j}{\sum_{j=1}^{94} e_j}$$

普通はおかしい、真の値からばらつきがあるはず

折衷案

$$\lambda_i = (1-\alpha)\frac{y_i}{e_i} + \alpha\frac{\sum_{j=1}^{94} y_j}{\sum_{j=1}^{94} e_j}$$

「死亡率は等しい」不具合の確認(1)

$$y_i \sim Po(e_i \lambda_i)$$

$$g(\lambda) \propto \frac{1}{\lambda}$$

真の死亡率の事前分布

$$g(\lambda | data) \propto \frac{1}{\lambda} \prod_{j=1}^{94} \lambda^{y_i} \exp(-e_j \lambda)$$

$$= \lambda^{\sum_{j=1}^{94} y_j - 1} \exp\left(-\sum_{j=1}^{94} e_j \lambda\right)$$

ガンマ分布

(参考) ポアソン分布

$$X \sim Po(\lambda) \quad \Rightarrow \quad p(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

$$g(\lambda|data) = \frac{g(\lambda)g(data|\lambda)}{g(data)} \propto g(\lambda)g(data|\lambda)$$

$$= \frac{1}{\lambda}\prod_{j=1}^{94}\frac{(e_j\lambda)^{y_j}e^{-e_j\lambda}}{y_j!}$$

$$\propto \frac{1}{\lambda}\prod_{j=1}^{94}\lambda^{y_j}e^{-e_j\lambda}$$

(参考) ガンマ分布

$$X \sim G(k, \theta) \quad \Rightarrow \quad f(x) = x^{k-1} \frac{\theta^k e^{-\theta x}}{\Gamma(k)}$$

$$k = \sum_{j=1}^{94} y_j \qquad \theta = \sum_{j=1}^{94} e_j$$

$$= 277 \qquad = 294681$$

とおくと

$$g(\lambda | data) \propto \lambda^{\sum_{j=1}^{94} y_j - 1} \exp\left(- \sum_{j=1}^{94} e_j \lambda \right)$$

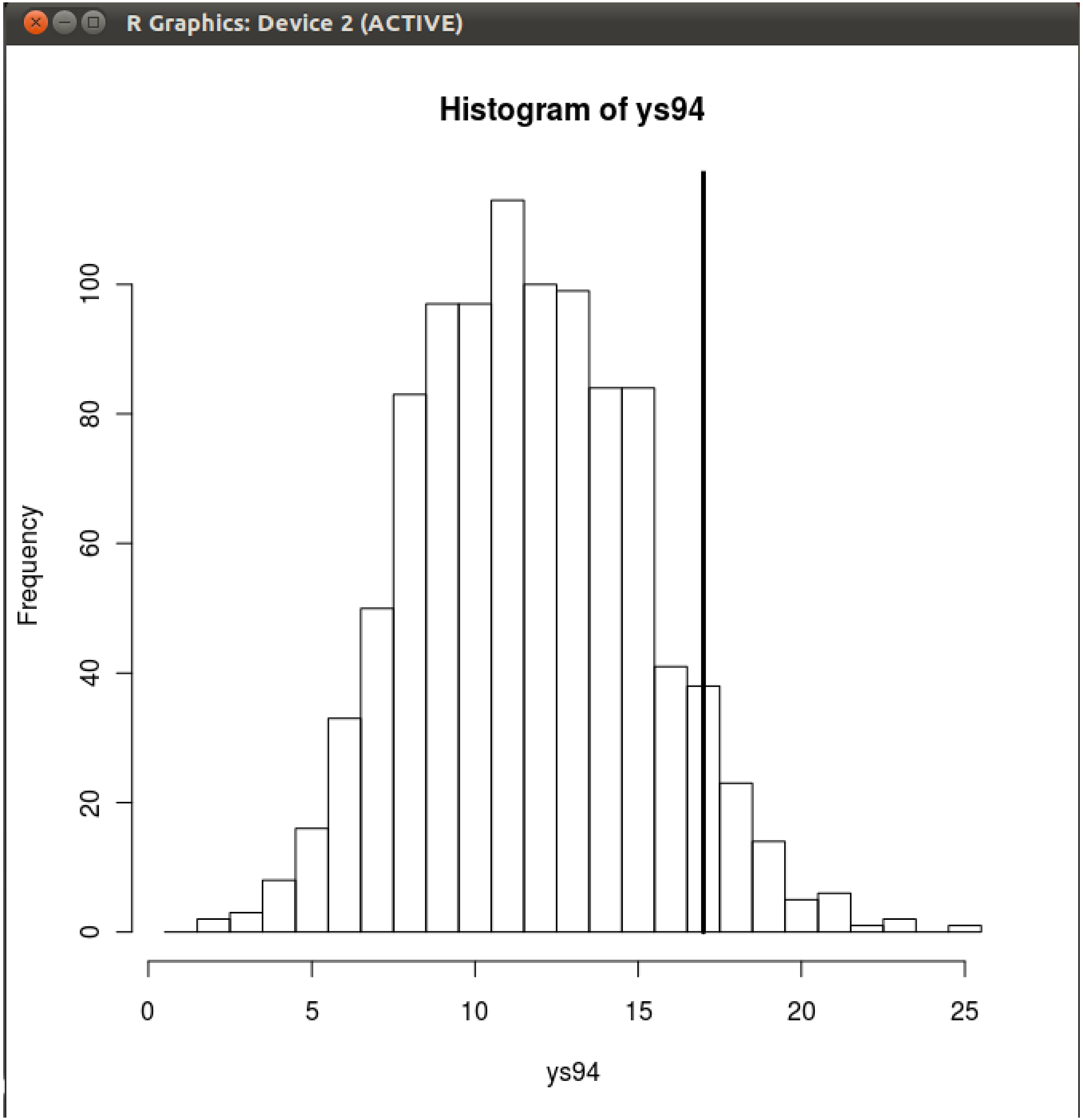
$$\propto f(\lambda)$$

「死亡率は等しい」不具合の確認(2)

94番目の病院の死亡数を先のモデルでシミュレーション

$y_{94} = 17$ ← 実際の値

```
> lambda <- rgamma(1000, shape = 277, rate = 294681)
> ys94 <- rpois(1000, e[94] * lambda)
> hist(ys94, breaks =seq(0.5, max(ys94) + 0.5))
> lines(c(y[94],y[94]), c(0,120), lwd = 3)
```


R Graphics: Device 2 (ACTIVE)
Histogram of ys94
Frequency
0
20
40
60
80
100
0
5
10
15
20
25
ys94

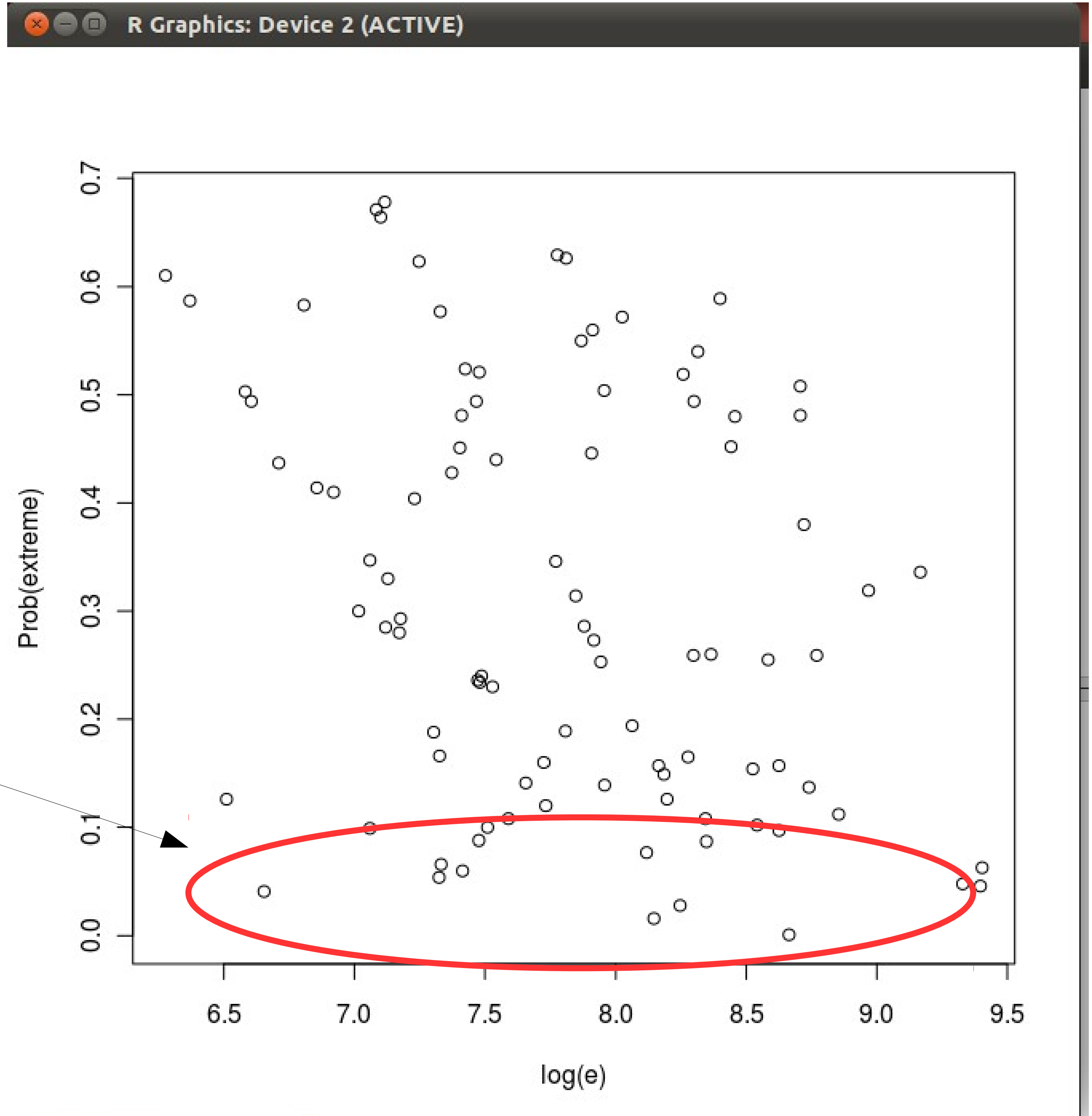
「死亡率は等しい」不具合の確認(3)

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

実際の死亡数と予測死亡数が離れている度合い

```
> lambda <- rgamma(1000, shape = 277, rate = 294681)
> pout <- sapply(1:94,prob.out)
> plot(log(e), pout, ylab="Prob(extreme)")
```

少ない


R Graphics: Device 2 (ACTIVE)
Prob(extreme)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
6.5
7.0
7.5
8.0
8.5
9.0
9.5
log(e)

交換可能性を事前の確信とするモデル

真の死亡率 λ_i は $G(\alpha, \alpha/\mu)$ からのサンプル

$$g(\lambda|\alpha, \mu) = \frac{(\alpha/\mu)^{\alpha}\lambda^{\alpha-1}e^{-\alpha\lambda/\mu}}{\Gamma(\alpha)}$$

$$\mu \sim G^{-1}(a, b) \qquad g(\mu) = \mu^{-a-1}e^{-b/\mu}$$

$$\alpha \sim g(\alpha)$$

逆ガンマ分布

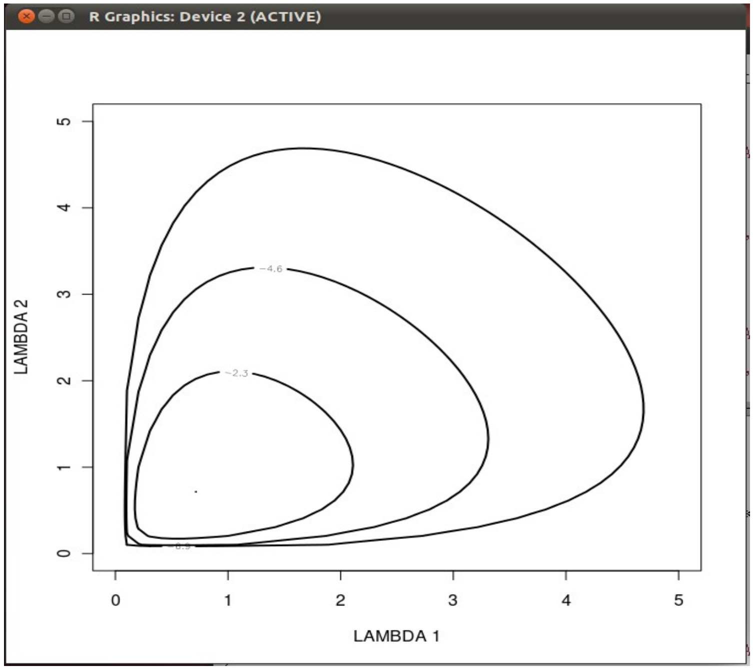
病院1と2 で確認

$$\mu \sim G^{-1}(a,b) \qquad \alpha \sim g(\alpha) = \alpha_0$$

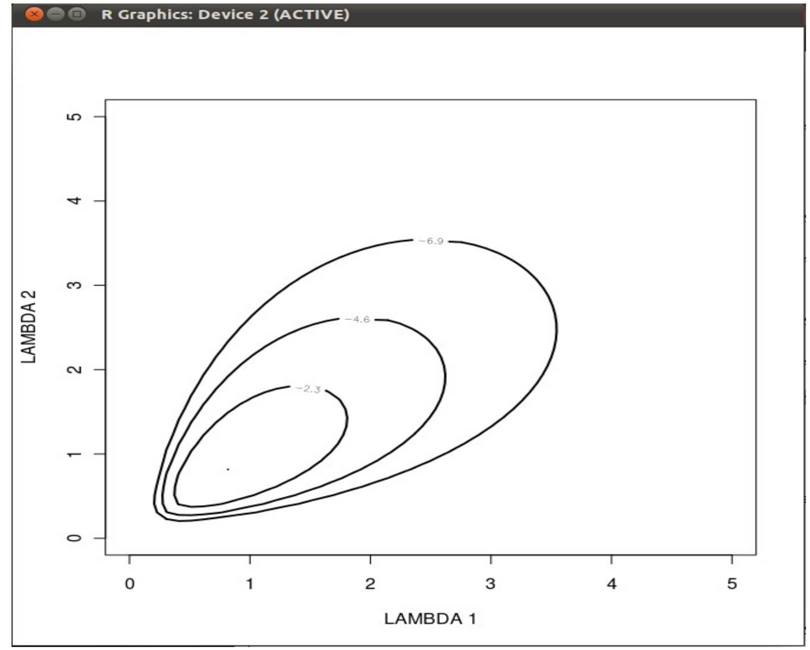
$$g(\lambda_1, \lambda_2 | \alpha_0) = \frac{(\lambda_1 \lambda_2)^{\alpha_0 - 1}}{(\alpha_0(\lambda_1 + \lambda_2) + b)^{2\alpha_0 + a}}$$

```
> alpha <- c(5,20, 80,400);
> for(j in 1:4) {
    mycontour(pgexchprior, c(.001, 5, .001, 5),
            c(alpha[j], 10, 10),
            main <- paste("ALPHA=", alpha[j]),
            xlab = "LAMBDA 1", ylab = "LAMBDA 2")
}
```

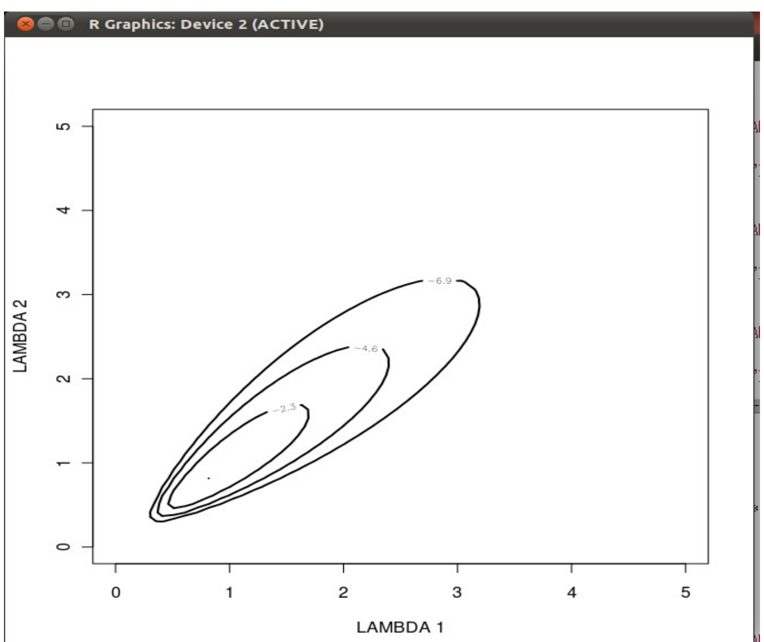
$\alpha = 5$



$\alpha = 20$



$\alpha = 80$



$\alpha = 400$