Hierarchical Models II

Statistics 220

Spring 2005

Computation in Hierarchical Models

If we want to do exact calculation of the posterior distribution, the following approach is useful

1. Joint posterior
$$p(\theta, \phi|y) \propto p(\phi)p(\theta|\phi)p(y|\theta, \phi)$$

2. Conditional posterior
$$p(\theta|\phi, y) \propto p(\theta|\phi)p(y|\theta, \phi)$$

   This will be easy if a conjugate prior is used.

3. Marginal posterior
$$p(\phi|y) = \int p(\theta, \phi|y)d\theta$$

Sometimes this can be gotten using the relationship

$$p(\phi|y) = \frac{p(\theta, \phi|y)}{p(\theta|\phi, y)}$$

Using this can be problematic, as the denominator has a normalizing 'constant' that depends on y and ϕ.

Often people will pick a prior $p(\phi)$ that is conjugate to $p(\theta|\phi)$, which will make this step easy. However, as we have seen, conjugate priors aren't always available.

This approach mirrors what is needed for direct simulation.

1. Sample $\phi_1, \ldots, \phi_m$ from $p(\phi|y)$
2. Sample $\theta_1, \ldots, \theta_m$ from $p(\theta|\phi_i, y)$
3. If necessary, sample $\tilde{y}$. The form of this draw depends on whether the θ of interest is one corresponding to the dataset or a new one.

Lets take a look at the rat tumors example

- Data model: $y_i =$ number of tumors in group i

$$y_i|\theta_i \overset{ind}{\sim} Bin(n_i, \theta_i) \quad i = 1, \ldots, 71$$

- Process model: $\theta_i =$ tumor rate in group i

$$\theta_i \overset{ind}{\sim} Beta(\alpha, \beta)$$

- Parameter model: $\alpha, \beta \sim p(\alpha, \beta)$.

The posterior distributions of interest are

- Joint posterior

$$
\begin{aligned}
p(\theta, \alpha, \beta | y) &\propto p(\alpha, \beta) p(\theta | \alpha, \beta) p(y | \theta, \alpha, \beta) \\
&\propto p(\alpha, \beta) \prod_{j=1}^{J} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta_j^{\alpha - 1} (1 - \theta)^{\beta - 1} \prod_{j=1}^{J} \theta_j^{y_j} (1 - \theta)^{n_j - y_j}
\end{aligned}
$$

Note that in this case we need the $\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$ terms in this (can't treat it as a constant and drop it) as α and β are random and have a prior on them.

- Conditional posterior

$$p(\theta|\alpha, \beta, y) = \prod_{j=1}^{J} \frac{\Gamma(\alpha + \beta + n_j)}{\Gamma(\alpha + y_j)\Gamma(\beta + n_j - y_j)} \theta_j^{\alpha + y_j - 1}(1 - \theta)^{\beta + n_j - y_j - 1}$$

So conditionally, the θ_j's are independent Beta's.

- Marginal posterior

Given the conjugate structure in the problem, integrating out the θ_j's is easy, giving

$$p(\alpha, \beta|y) \propto p(\alpha, \beta) \prod_{j=1}^{J} \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)} \frac{\Gamma(\alpha + \beta + n_j)}{\Gamma(\alpha + y_j)\Gamma(\beta + n_j - y_j)}$$

This does not have a simpler form, but the density can be calculated easily up to the normalizing constant and can be simulated from.

In the book, they put following vague prior on α and β

$$p(\alpha, \beta) \propto \frac{1}{(\alpha + \beta)^{5/2}}$$

which comes from putting a uniform prior on

$$\frac{\alpha}{\alpha + \beta}, \frac{1}{\sqrt{\alpha + \beta}}$$

Note that this prior puts large weight on α and β both being small.

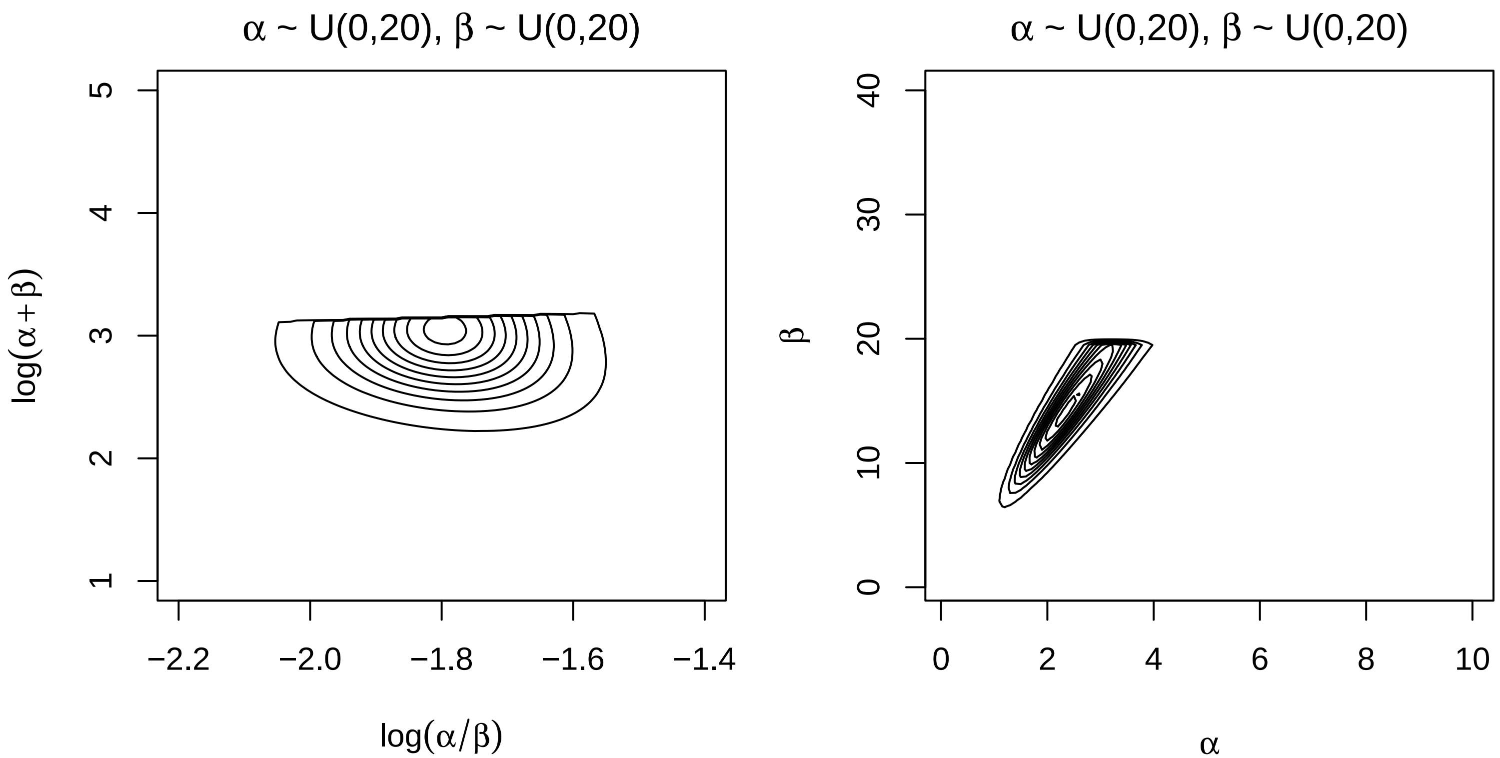
Instead, of this, lets put the following independent prior on α and β

$$\alpha \sim Unif(0, 20) \qquad \beta \sim Unif(0, 20)$$

(this is the same prior used in the previous day's notes)

This gives

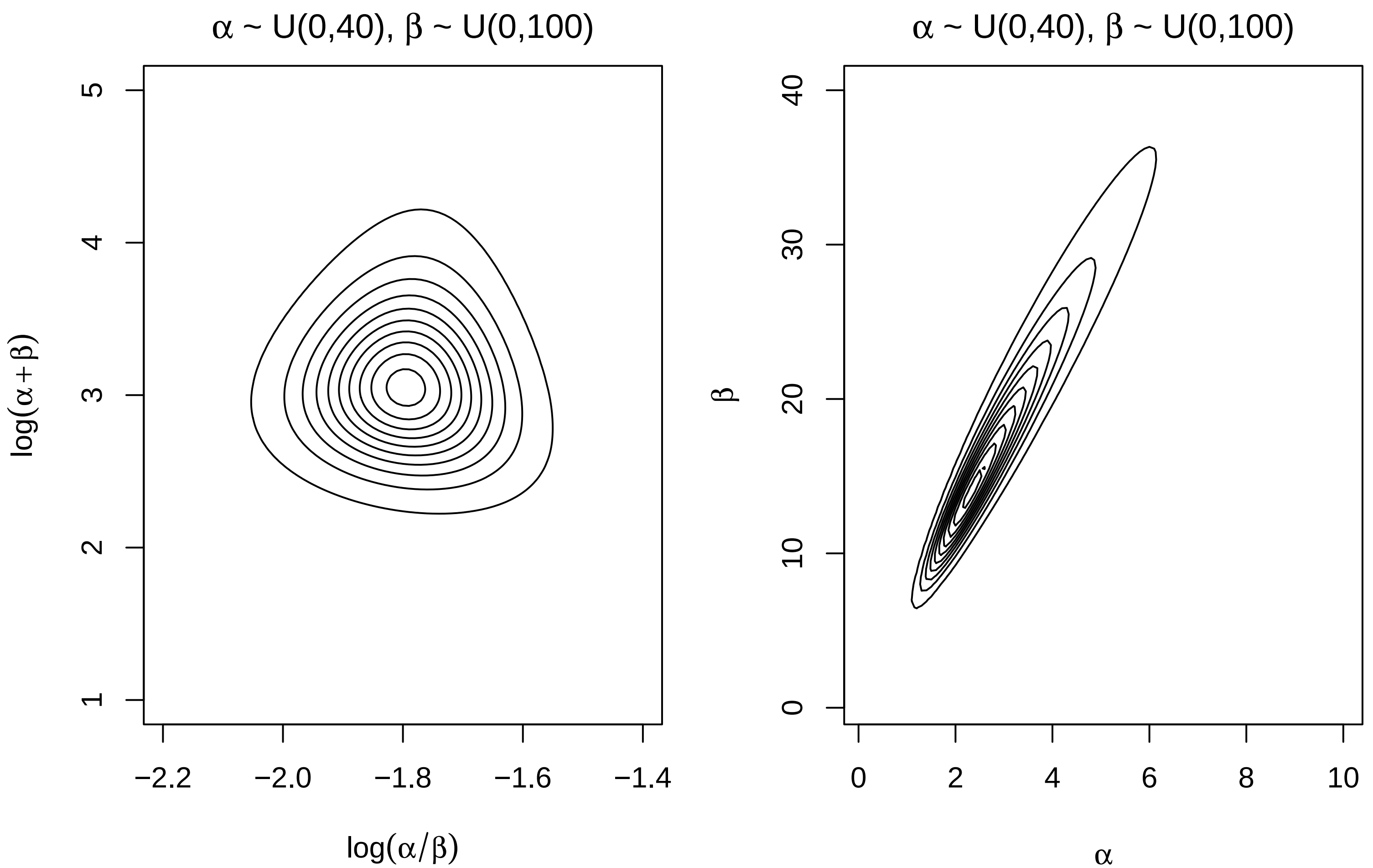
$$p(\alpha, \beta|y) \propto I(\alpha \leq 20)I(\beta \leq 20) \prod_{j=1}^{J} \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \frac{\Gamma(\alpha+\beta+n_j)}{\Gamma(\alpha+y_j)\Gamma(\beta+n_j-y_j)}$$



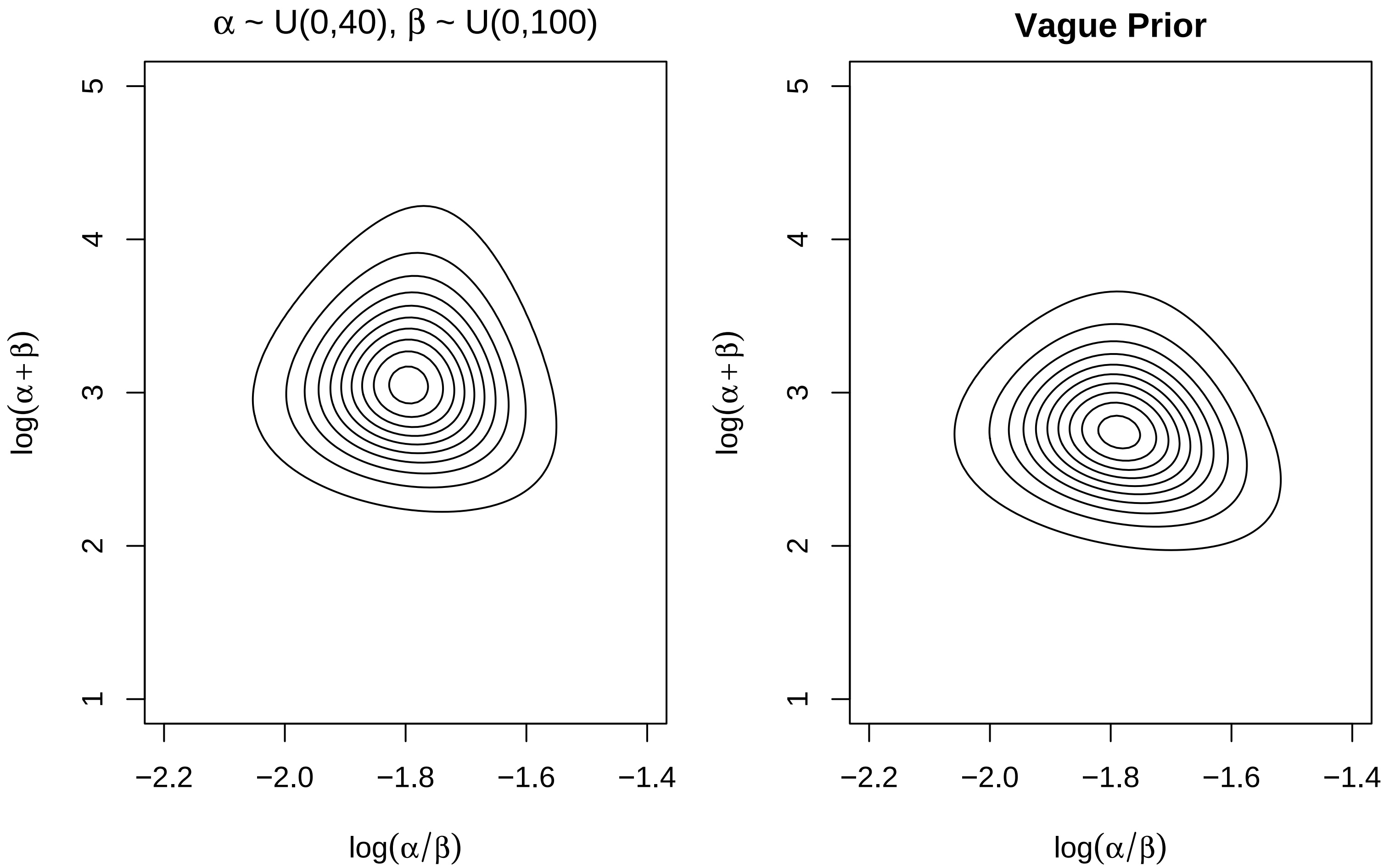
Note that the posterior gets clipped due to the upper limits on α and β.

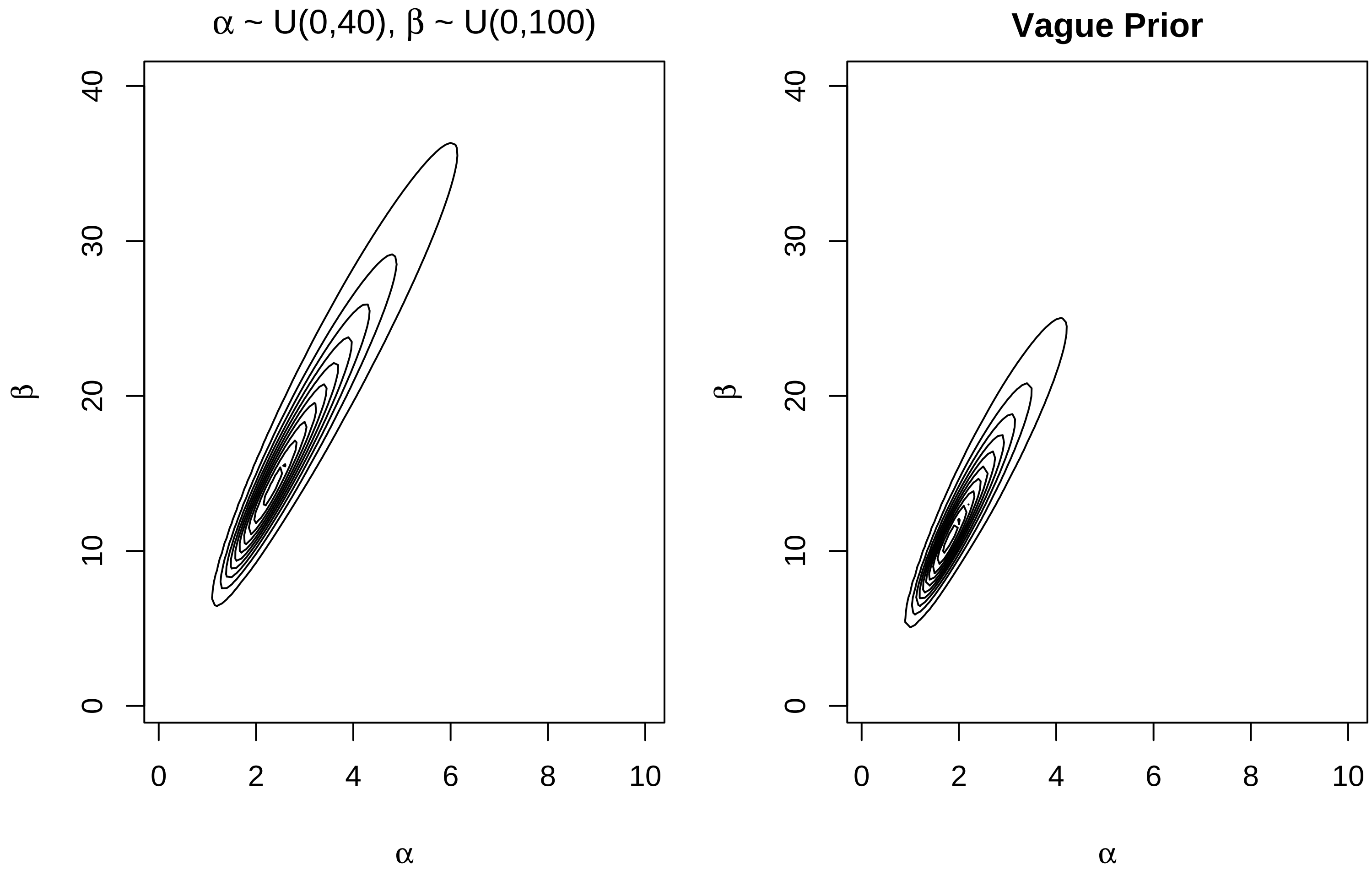
So naive implementation of uniform priors can be highly informative.

Lets extend those limits so the choices match the data better.



Now lets compare the marginal posterior under this prior with the posterior under the vague prior suggested by the book.





So as expected, the vague prior pulls $\alpha + \beta$ down.

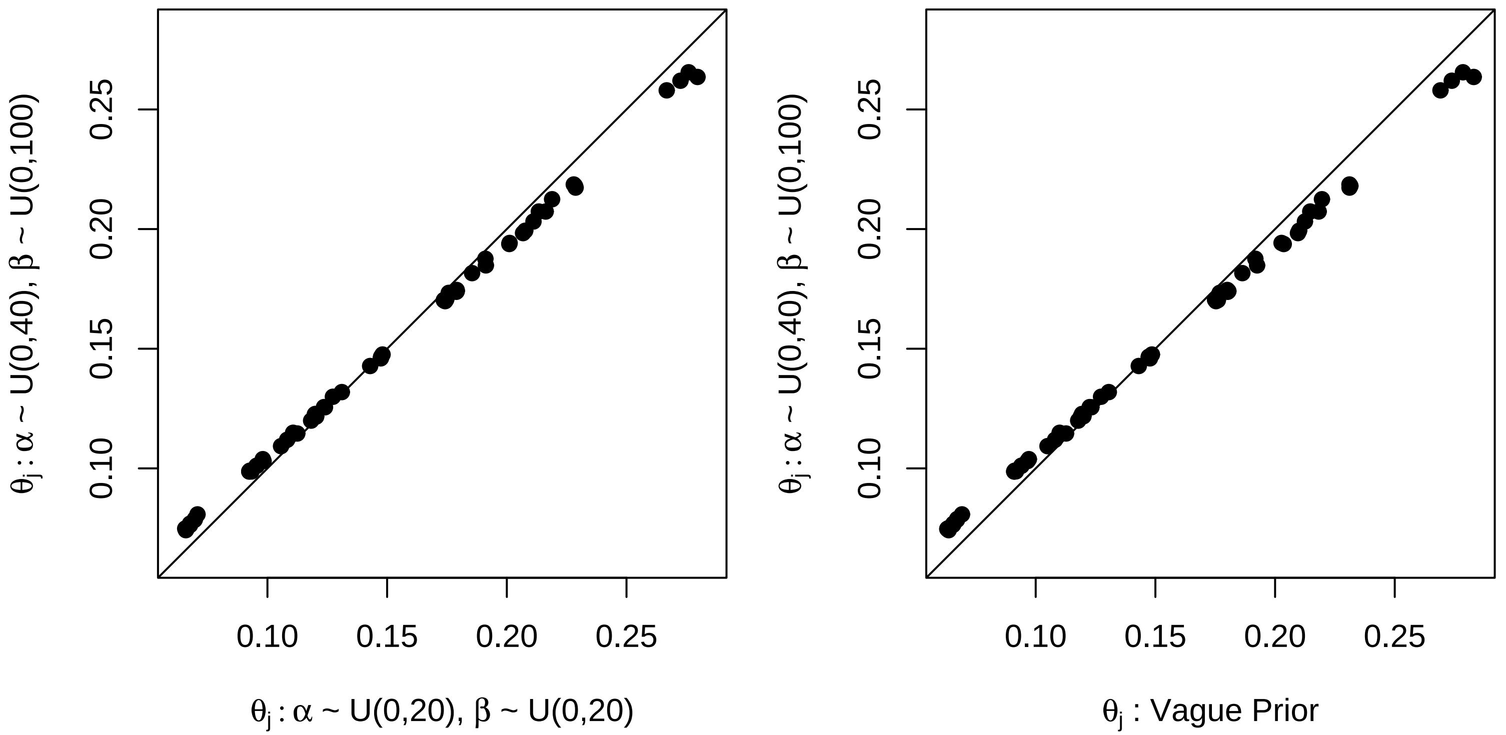
The posterior means of α and β (as calculated by simulation) are

Prior	Vague	$\alpha \sim U(0,20), \beta \sim U(0,20)$	$\alpha \sim U(0,40), \beta \sim U(0,100)$
α	2.398	2.482	3.448
β	14.291	14.805	20.649

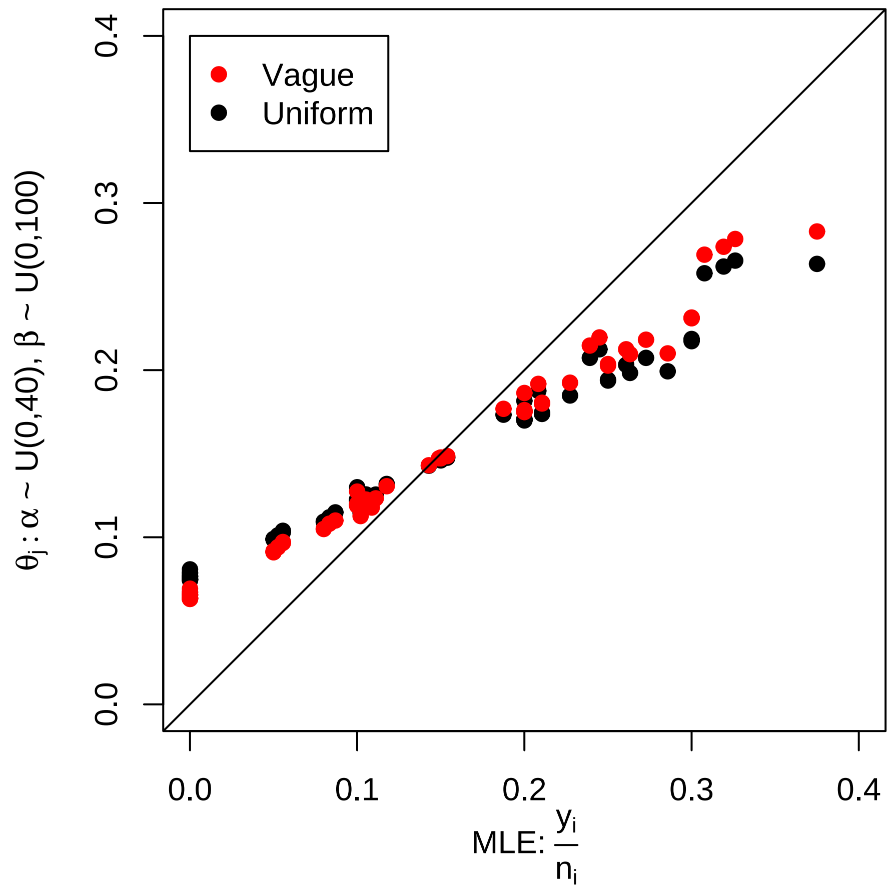
Now we are really interested in the θ_j, the tumor rates in the different groups. So we want the determine

$$p(\theta|y) = \int p(\theta|\alpha,\beta,y)p(\alpha,\beta|y)d\alpha d\beta$$

This does not have a nice closed form as integrating α and β is ugly, so we will have to use simulation.



0.10
0.15
0.20
0.25
θj : α ~ U(0,20), β ~ U(0,20)
θj : α ~ U(0,40), β ~ U(0,100)
θj : Vague Prior



In this case the $\alpha \sim Unif(0, 40), \beta \sim Unif(0, 100)$ prior shrinks the estimates more than the vague prior, though they are shrinking to about the same place.

This is supported by the posterior means (as calculated by simulation) for

Prior	Vague	$\alpha \sim Unif(0, 40), \beta \sim Unif(0, 100)$
$\frac{\alpha}{\alpha+\beta}$	0.144	0.143
$\alpha + \beta$	16.689	24.097

For a new group

$$E[\theta|y] = E[E[\theta|\alpha, \beta, y]] = E\left[\frac{\alpha}{\alpha + \beta}|y\right]$$

so the two priors seem to be shrinking to the same place.

For an observed group j,

$$E[\theta_j|y] = E[E[\theta_j|\alpha,\beta,y]] = E\left[\frac{\alpha + y_j}{\alpha+\beta+n_j}|y\right]$$

Note that

$$\frac{\alpha + y_j}{\alpha+\beta+n_j} = \frac{\alpha+\beta}{\alpha+\beta+n_j}\frac{\alpha}{\alpha+\beta} + \frac{n}{\alpha+\beta+n_j}\frac{y_j}{n_j}$$

so this agrees with more shrinking for the uniform prior as the effective sample size from the prior component $(\alpha+\beta)$ is larger for the uniform prior.