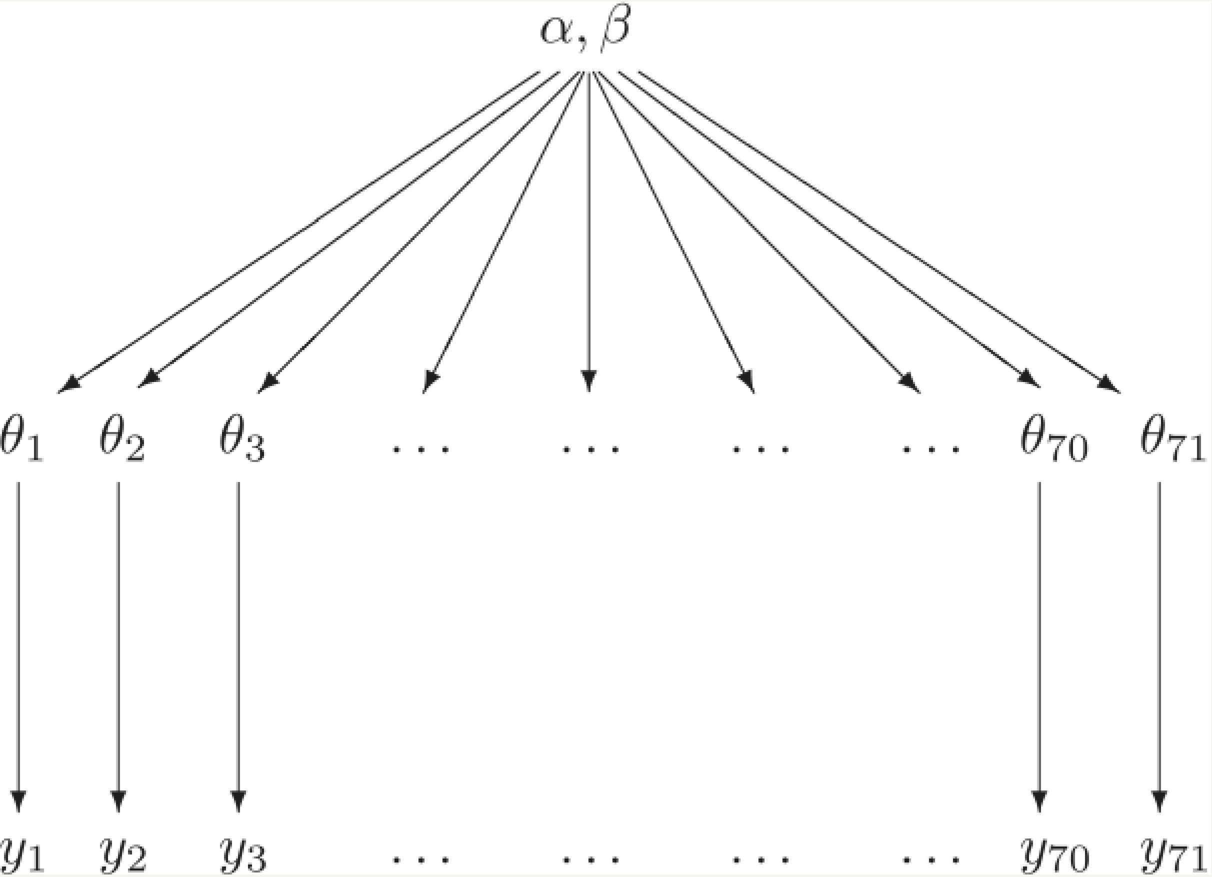
Hierarchical beta-binomial model

Stat 340: Bayesian Statistics

Motivating experiment

- Imagine a single toxicity experiment performed on rats. Lots of those are carried out before drugs are approved for use in humans.
- θ is the probability that a rat receiving no treatment in an experiment develops a tumor.
- Current experiment: $n = 14$ rats in the study, and $y = 4$ rats develop a tumor.
- Previous experiments: the same experiment has been performed with 70 other groups of rats
  - $y_j =$ # rats with tumors in jth experiment
  - $n_j =$ sample size in jth experiment
  - $\theta_j =$ probability that a rat receiving no treatment in an experiment develops a tumor in jth experiment

Structure of hierarchical model



Assumption:

current tumor risk, θ_{71}, and the 70 historical risks, $\theta_1, \ldots, \theta_{70}$, are a random sample from a common distribution

Fig. 5.1 in Gelman et al. (2004)

Posterior derivations

Before moving on to hyper-prior selection and tuning, let's explore how to derive out posterior distributions

Hyper priors (second stage)

- In hierarchical models, the choice of the hyper-prior is important because it is possible to end up with an improper posterior distribution.
- For example, in this beta-binomial example a uniform prior on $\alpha, \ \beta$ does not work because the posterior distribution of $\alpha, \ \beta$ is non-integrable.
- An approach that often makes the choice of hyper prior easier is to **think of functions of the parameters that are more intuitive.**

Beta-binomial hyperprior

What are functions of α and β that are more intuitive?

Prior mean: $$\mu = \frac{\alpha}{\alpha + \beta}$$

Prior sample size: $$\eta = \alpha + \beta$$

Approx. prior SD: $$\eta^* = (\alpha + \beta)^{-1/2}$$

Noninformative hyperprior - option 1

One way to develop a noninformative prior would be to place uniform densities on (μ, η)

$$\mu = \frac{\alpha}{\alpha + \beta} \sim \mathrm{Beta}(1, 1)$$
$$\eta = \alpha + \beta \sim 1$$

What does this imply about the prior of α, β on the original scale?

We need to find the Jacobian of the inverse transformation to answer this.

Noninformative hyperprior - option 2

An alternative approach is to place uniform densities on prior the mean and approximate SD

$$\mu = \frac{\alpha}{\alpha + \beta} \sim \mathrm{Beta}(1, 1)$$
$$\eta^* = (\alpha + \beta)^{-1/2} \sim 1$$

The implied prior for (α, β) is given by

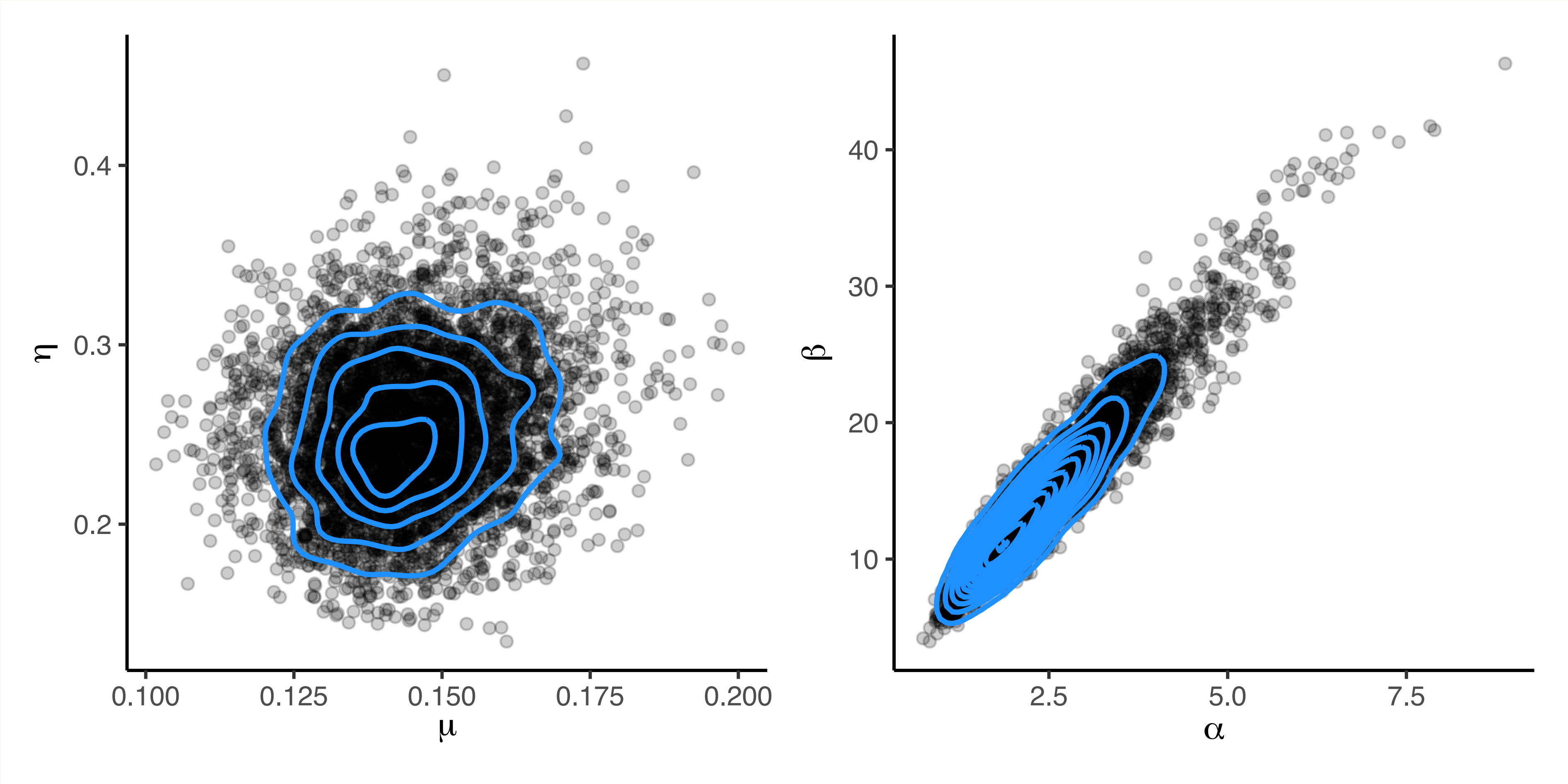
$$\pi(\alpha, \beta) \propto (\alpha + \beta)^{-5/2}$$

which is a proper prior distribution (should be more efficient)

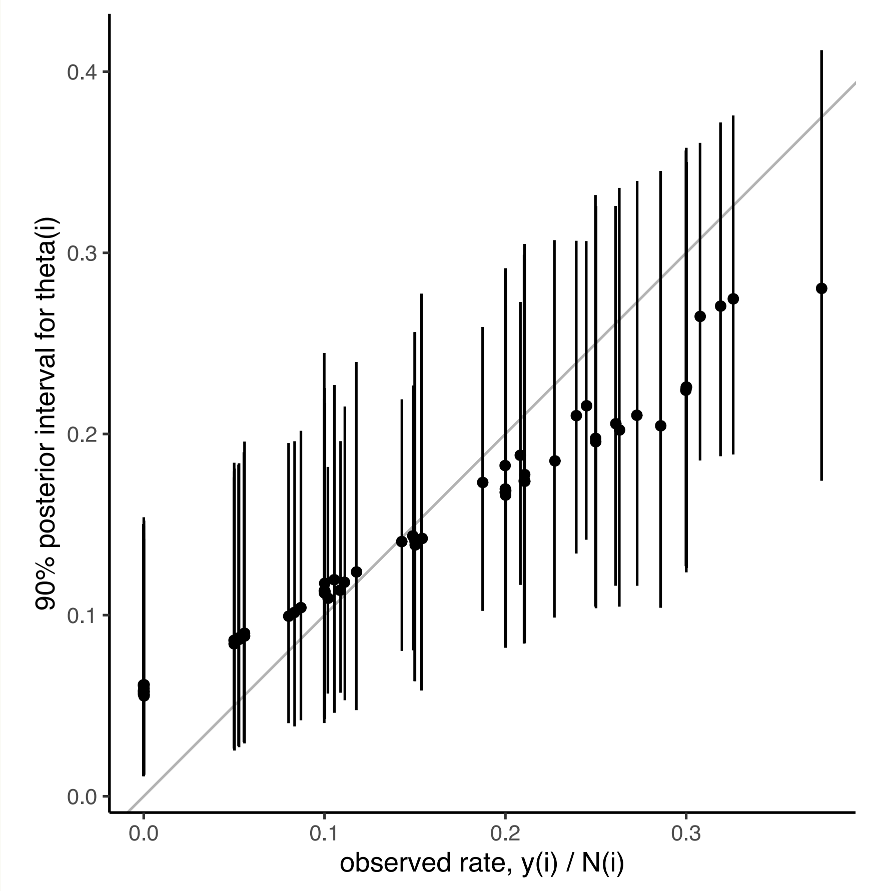
Noninformative priors in JAGS

```
noninform_model<-"
model {
## sampling
for (i in 1:N){
   y[i] ~ dbin(theta[i], n[i])
}

## priors
for (i in 1:N){
   theta[i] ~ dbeta(alpha, beta)
}

## noninformative hyperpriors
alpha <- mu / pow(eta, 2)
beta  <- (1 - mu) / pow(eta, 2)
mu ~ dbeta(1, 1)
eta ~ dbeta(1, 1)
}"
```

Comparing posterior parameterizations



Estimated vs observed tumor rates



- The rates θ_j are **shrunk** toward their sample point estimates (diagonal)
- Smaller experiments are shrunk more and have higher posterior variances

What if we don't want to be "fully noninformative"

Hyperprior for μ

- $\mu = \dfrac{\alpha}{\alpha + \beta} \in (0, 1)$
- $\mu \sim \mathrm{Beta}(a_0, b_0)$ is reasonable
- Beta(1, 1) would represent little prior knowledge

Hyperprior for η

- $\eta > 0$
- Many options for distributions, which makes sense?

Albert and Hu's approach

- Choose $\mu \sim \mathrm{Beta}(a_0, b_0)$
- Reframe η in terms of a shrinkage factor, λ, and place a prior on λ
  - $\theta_j|\alpha, \beta, y \sim \mathrm{Beta}(\alpha + y_j, \beta + n_j - y_j)$
  - $E(\theta_j|\alpha, \beta, y) = \dfrac{\alpha + y_j}{n_j + \alpha + \beta}$, now re-express in terms of μ and η

Albert and Hu's approach

- Once you tune your prior on λ, this induces a prior on η

$$\lambda \sim \text{Unif}(0,1) \qquad \Longrightarrow \qquad \pi(\eta) = \frac{n_j}{(n_j+\eta)^2}, \eta > 0$$

- JAGS doesn't "know" this distribution, but it knows the **logistic distribution**
- Let $\nu = \log(\eta)$, this transformation gives a $\text{Logistic}(n_j, 1)$ PDF

$$\pi(\nu) = \frac{e^{-(\nu - \log n_j)}}{\left[1 + e^{-(\nu - \log n_j)}\right]^2}, \nu \in \mathbb{R}$$

- `dlogis` in JAGS

You can derive using the univariate transformation you learned in probability

JAGS model specification

```
weak_inform_model<-"
model {
## sampling
for (i in 1:N){
   y[i] ~ dbin(theta[i], n[i])
}

## priors
for (i in 1:N){
   theta[i] ~ dbeta(alpha, beta)
}

## noninformative hyperpriors
alpha <- mu * eta
beta  <- (1 - mu) * eta
mu ~ dbeta(1, 1)
eta <- exp(logeta)
logeta ~ dlogis(logn, 1)
}"
```

Comparing posterior parameterizations