ABDA

Chapter 9 Summary

by

Emad Ibrahim

Hierarchical models:

A model of **multiple parameters** such that the credible values of some parameters **depend** on the values of **other parameters**.

Example 1: coins minted from the same factory.

$$\theta_s(bias\ of\ single\ coin)\alpha\ \omega\ (bias\ of\ the\ factory)$$

Example 2: Batting ability of baseball players who have different field positions

θ_s(Each player's ability)$\alpha\ \omega$ (the typical ability of players in that position)

Example 3: the probability that a child bought lunch from the school cafeteria

θ_s(Each child's probability)$\alpha\ \omega$ (the typical buying probability of children in the school.)

Hierarchical models:

- The estimate of each individual parameter (θ_s) is simultaneously informed by data from all the other individuals($\theta_1, \theta_2, \ldots\ldots$).

- Because the higher-level parameters ,ω, inform from all individuals ($\theta_1, \theta_2, \ldots\ldots$).

- In hierarchical models the likelihood can be factored into a chain of dependencies:

$$\begin{aligned} p(\theta, \omega|D) &\propto p(D|\theta, \omega)\, p(\theta, \omega) \\ &= p(D|\theta)\, p(\theta|\omega)\, p(\omega) \end{aligned}$$

- The data are **independent of all other parameter** values when the value θ is set.

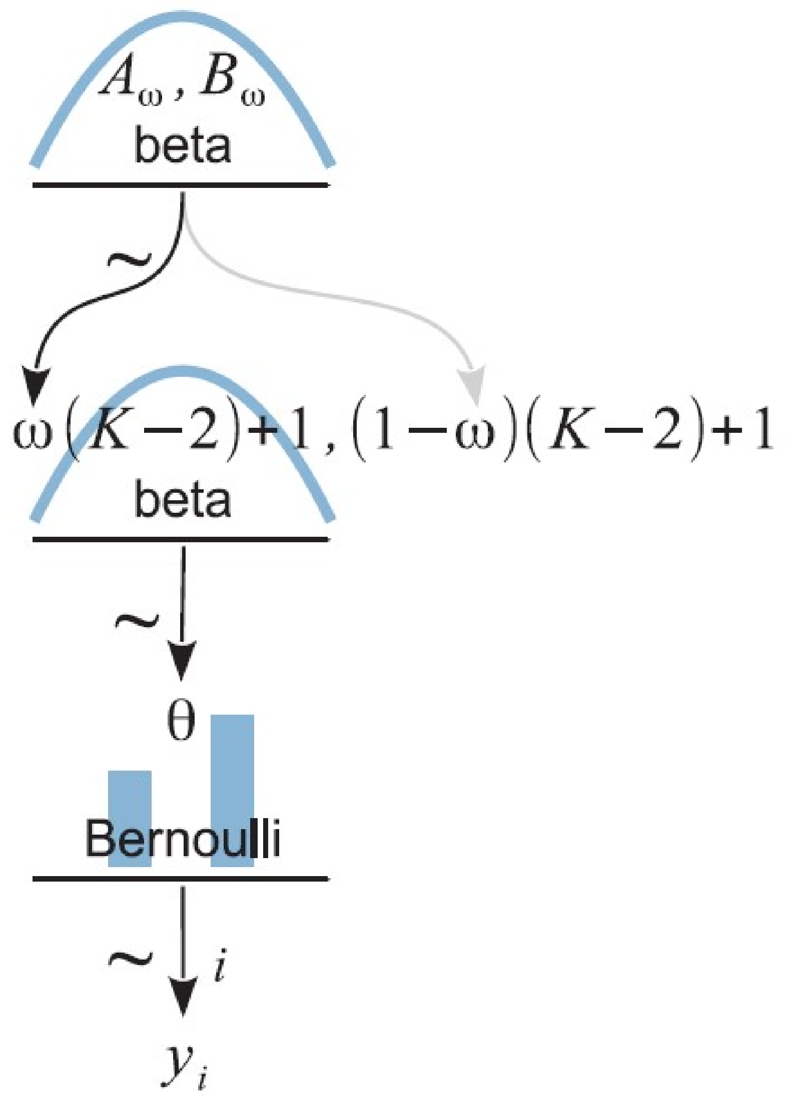
9.1. A SINGLE COIN FROM A SINGLE MINT:

- likelihood function is the Bernoulli Distribution and prior distribution is a beta density.

$$y_i \sim \text{dbern}(\theta) \qquad\qquad \theta \sim \text{dbeta}(a, b)$$

- The beta density, *a* and *b*, can be re-expressed in terms of the mode ω and concentration κ of the beta distribution.

$$\theta \sim \text{dbeta}\,(\,\omega(\kappa-2)+1\,,\;(1-\omega)(\kappa-2)+1\,)$$



- Think of ω **another parameter to be estimated** given the data.

- Let prior distribution on ω is a beta distribution:

$$p(\omega) = \text{beta}(\omega | A_\omega, B_\omega)$$

- The figure shows the dependency structure among the variables.
- $\omega \rightarrow \theta$, $\theta \rightarrow y$.

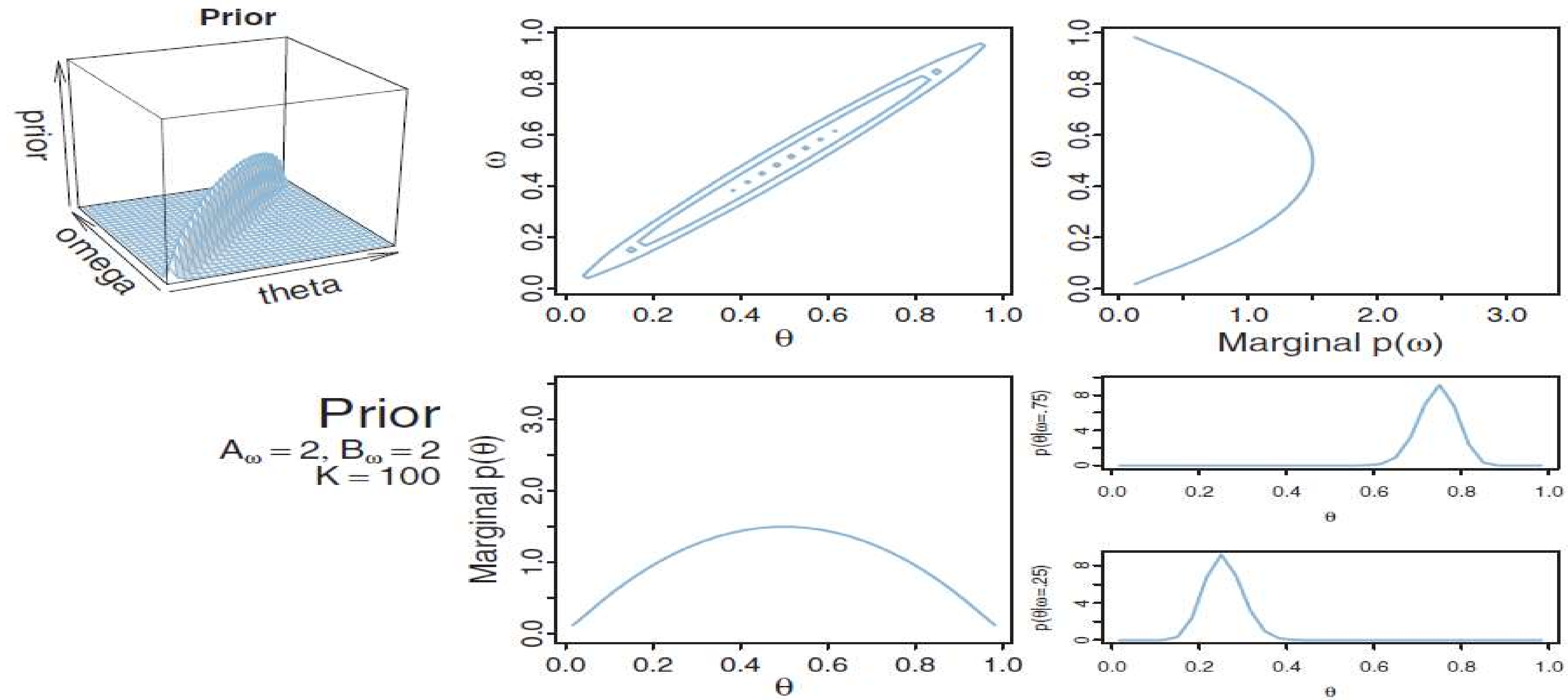
9.1. A SINGLE COIN FROMA SINGLE MINT:

- The posterior of two parameter (θ, ω) is:

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

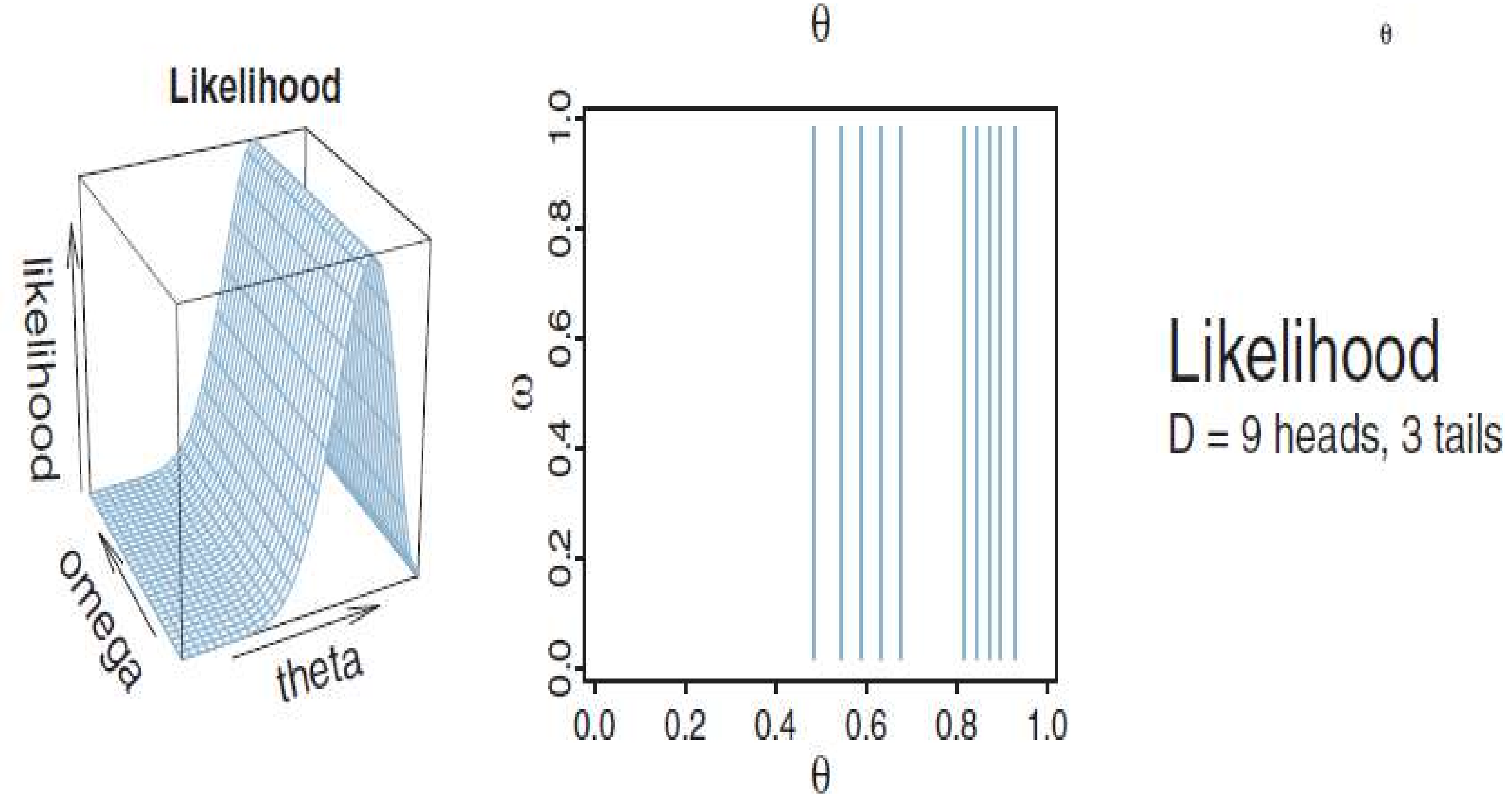
- *Let* A_ω = 2, B_ω = 2, *K* = 100. Then joint prior pdf:

$p(\theta, \omega) = p(\theta\,|\omega)p(\omega) = \text{beta}(\theta\,|\omega(100-2)+1, (1-\omega)(100-2)+1)\ \text{beta}(\omega|2, 2)$



9.1. A SINGLE COIN FROMA SINGLE MINT:

$p(\{y\}|\theta, \omega)=p(\{y\}|\theta)$ → likelihood function depends only on θ and not on ω.



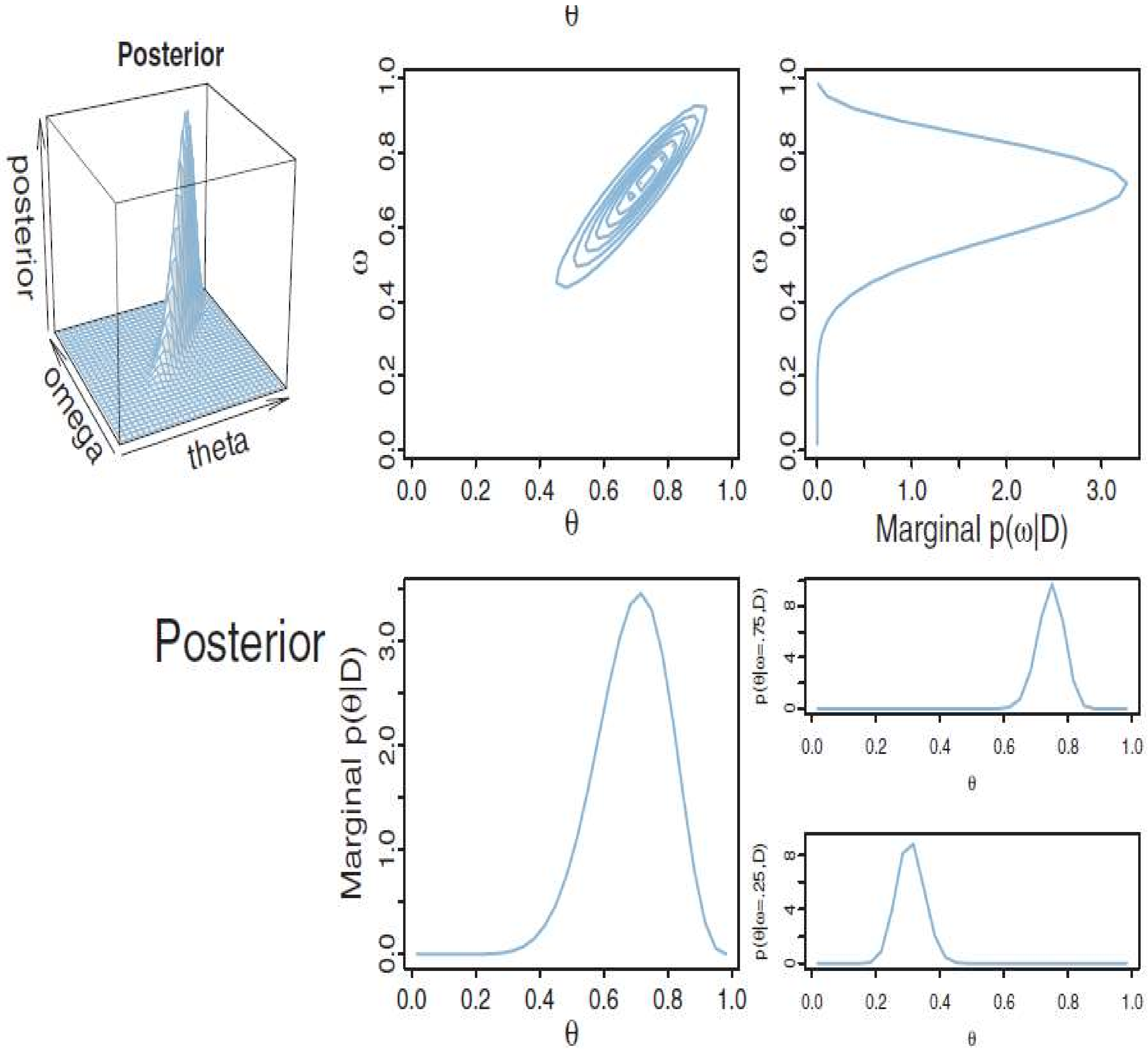
9.1. A SINGLE COIN FROMA SINGLE MINT:

p(θ |*ω*) ≈ *p*(θ |*ω*, {*yi*}).

- because the prior beliefs regarding the dependency of θ on *ω* had little uncertainty. K=100

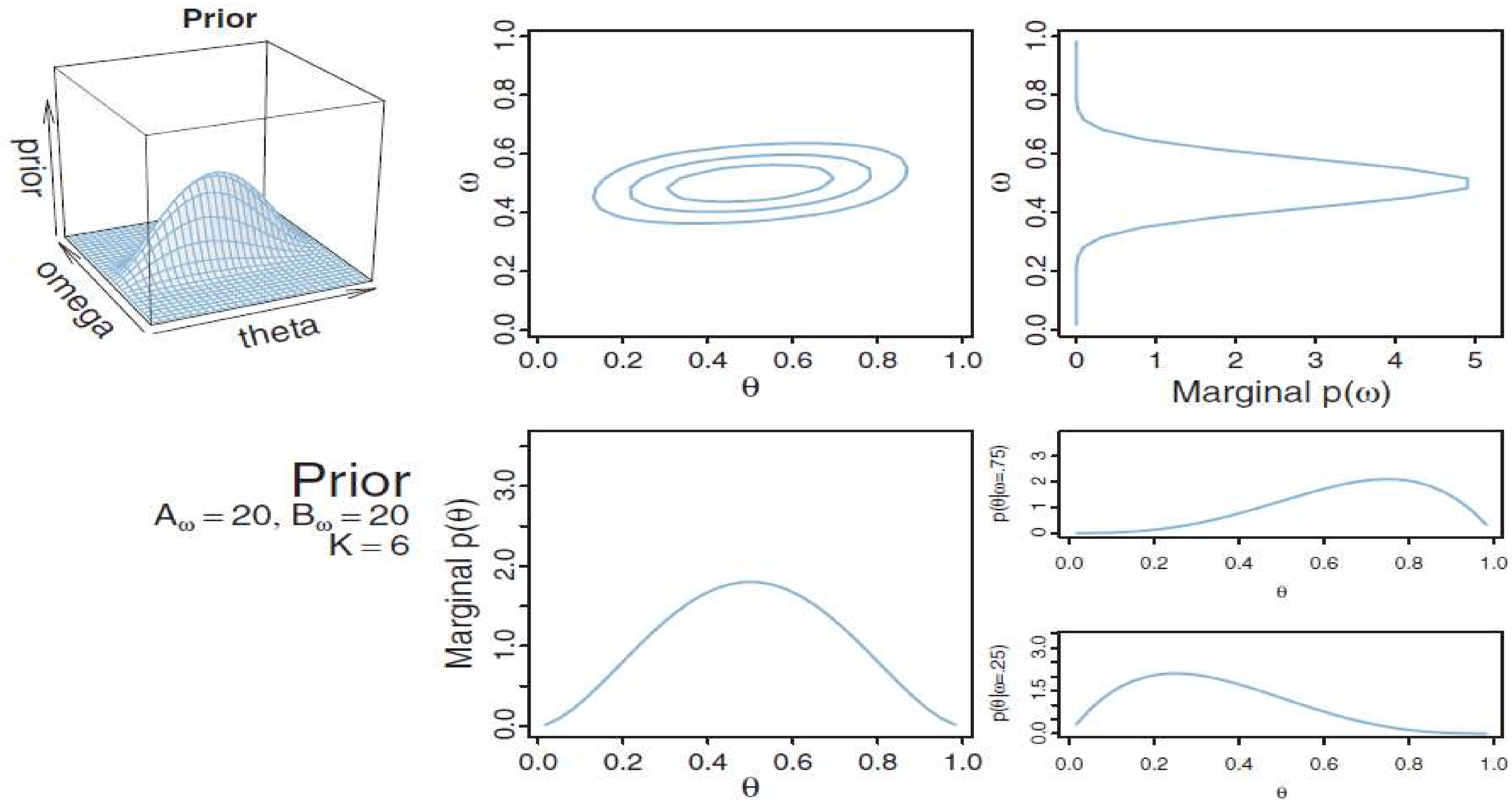
p(*ω*) and *p*(*ω*|{*yi*}) are very different.

- *p*(*ω*) very uncertain and therefore easily influenced by the data. A_ω = 2, B_ω = 2.



9.1. A SINGLE COIN FROMA SINGLE MINT:

- As a contrasting case, A_{ω} = 20, B_{ω} = 20, and K = 6.
- So, high prior certainty regarding ω, but low prior certainty regarding the dependence of θ on ω.
- *p(ω)* is fairly sharply peaked over ω =.5., *p(θ |ω)* are very broad.



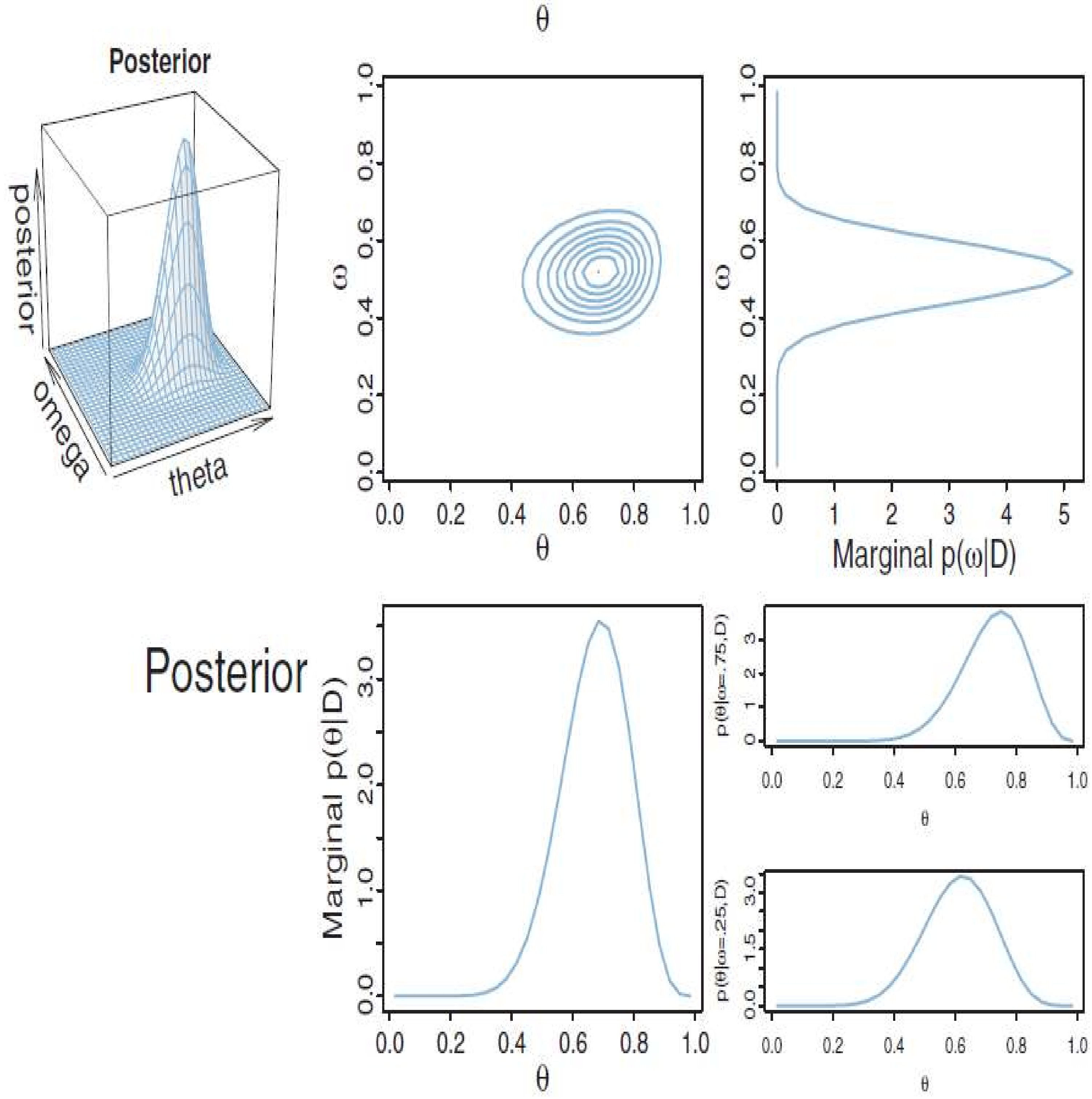
<u>9.1. A SINGLE COIN FROMA SINGLE MINT:</u>

*p(*θ *|ω)* and *p(*θ *|ω, {yi})* are very different.

- because the prior beliefs regarding the dependency of θ on *ω* had high uncertainty. K=6.

p(ω) ≈ *p(ω|{yi})*

- *p(ω)* has high certainty → A_ω = 20, B_ω = 20.



9.2. MULTIPLE COINS FROM A SINGLE MINT:

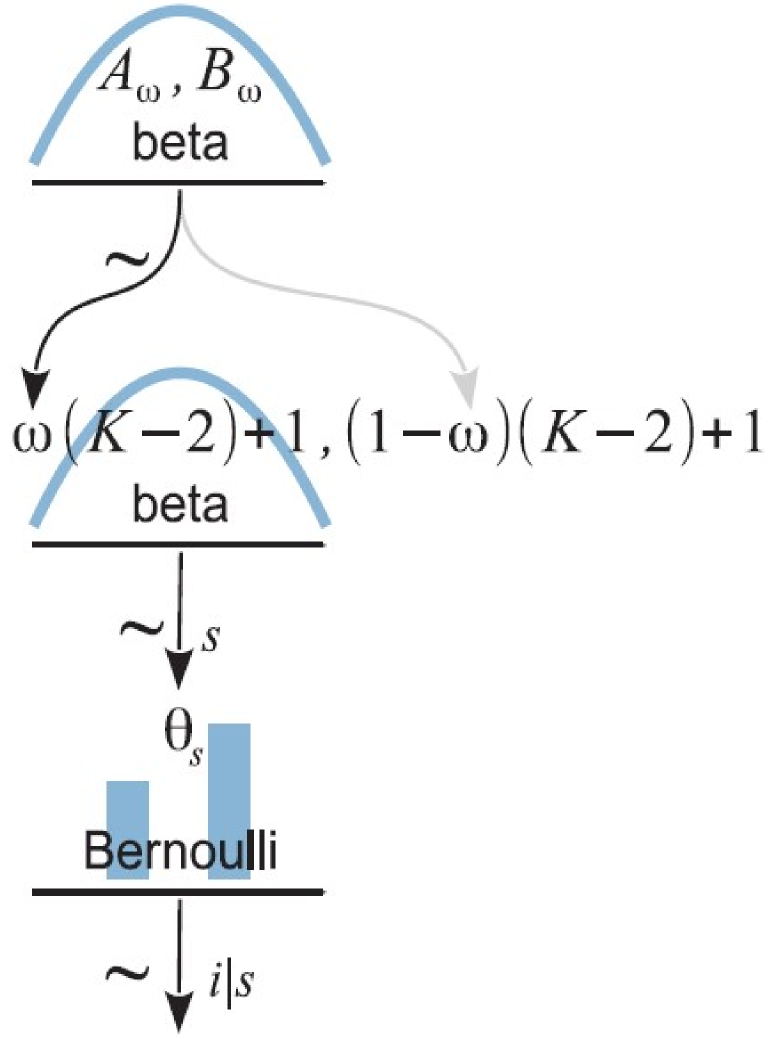
- Coin 1→ θ_1, Coin 2→ θ_2 …..Coin S→θ_s and the mode of the mint is ω.
- S+1 parameters.

- $P(\theta_s)$ is a Beta function:

$$\theta_s \sim \text{dbeta}(\omega(K-2)+1, (1-\omega)(K-2)+1)$$

- The likelihood is a Bernoulli distribution.

$$y_{i|s} \sim \text{dbern}(\theta_s)$$



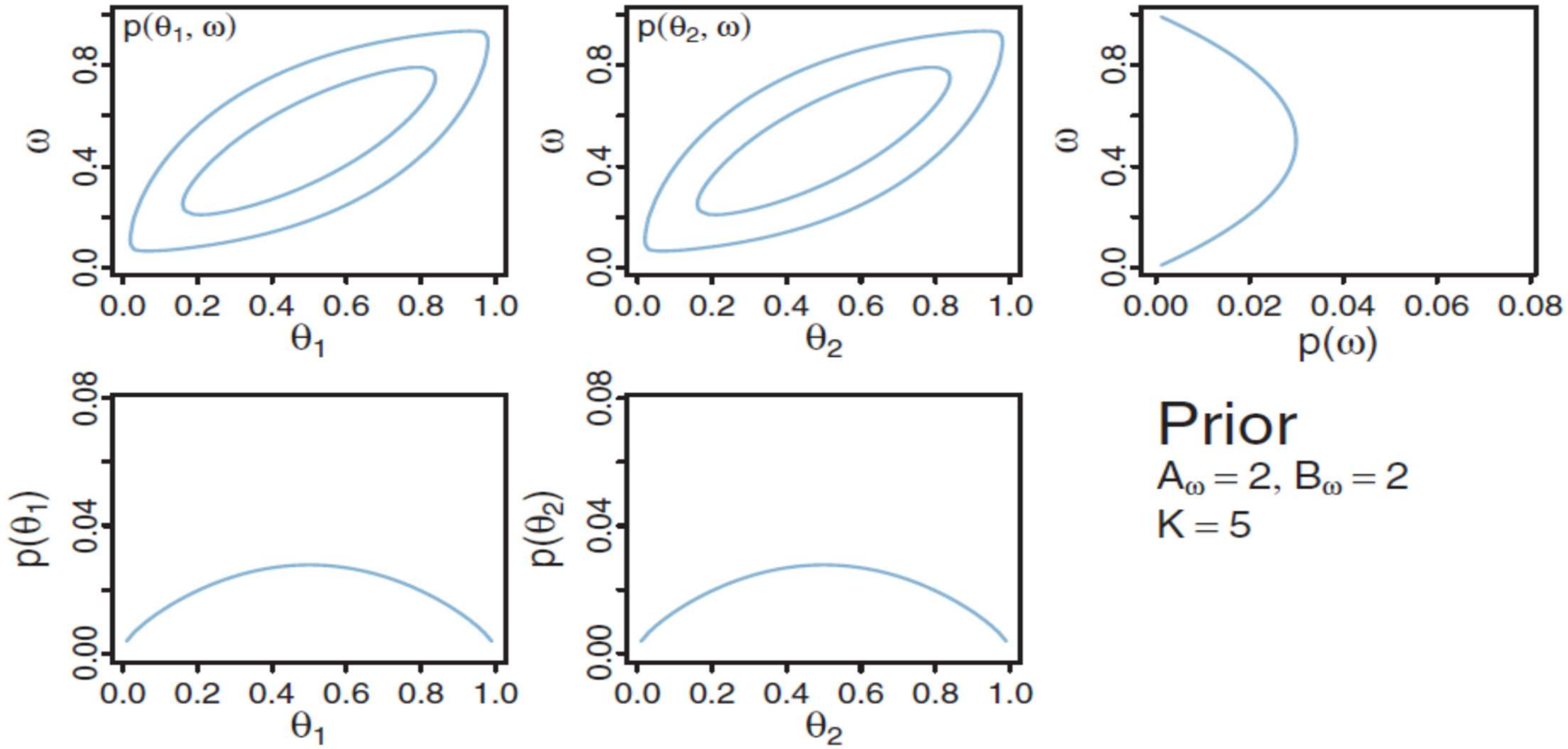
9.2. MULTIPLE COINS FROM A SINGLE MINT:

Example: Two coins

- Three parameters: θ_1, θ_2 and ω.
- *Let* A_ω = 2, B_ω = 2, *K* =5.

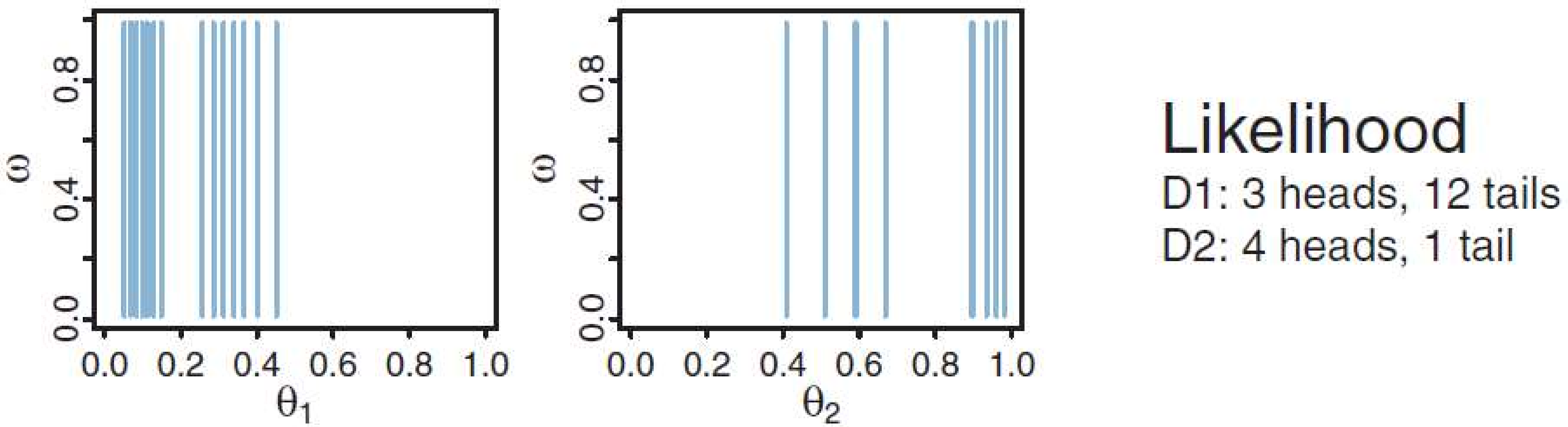
$$p(\theta_j|\omega) = \text{beta}(\theta_j|\omega(5-2)+1, (1-\omega)(5-2)+1);$$

- Then joint prior pdf: $p(\theta_1, \theta_2, \omega) = p(\theta_1|\omega)\, p(\theta_2|\omega)\, p(\omega)$



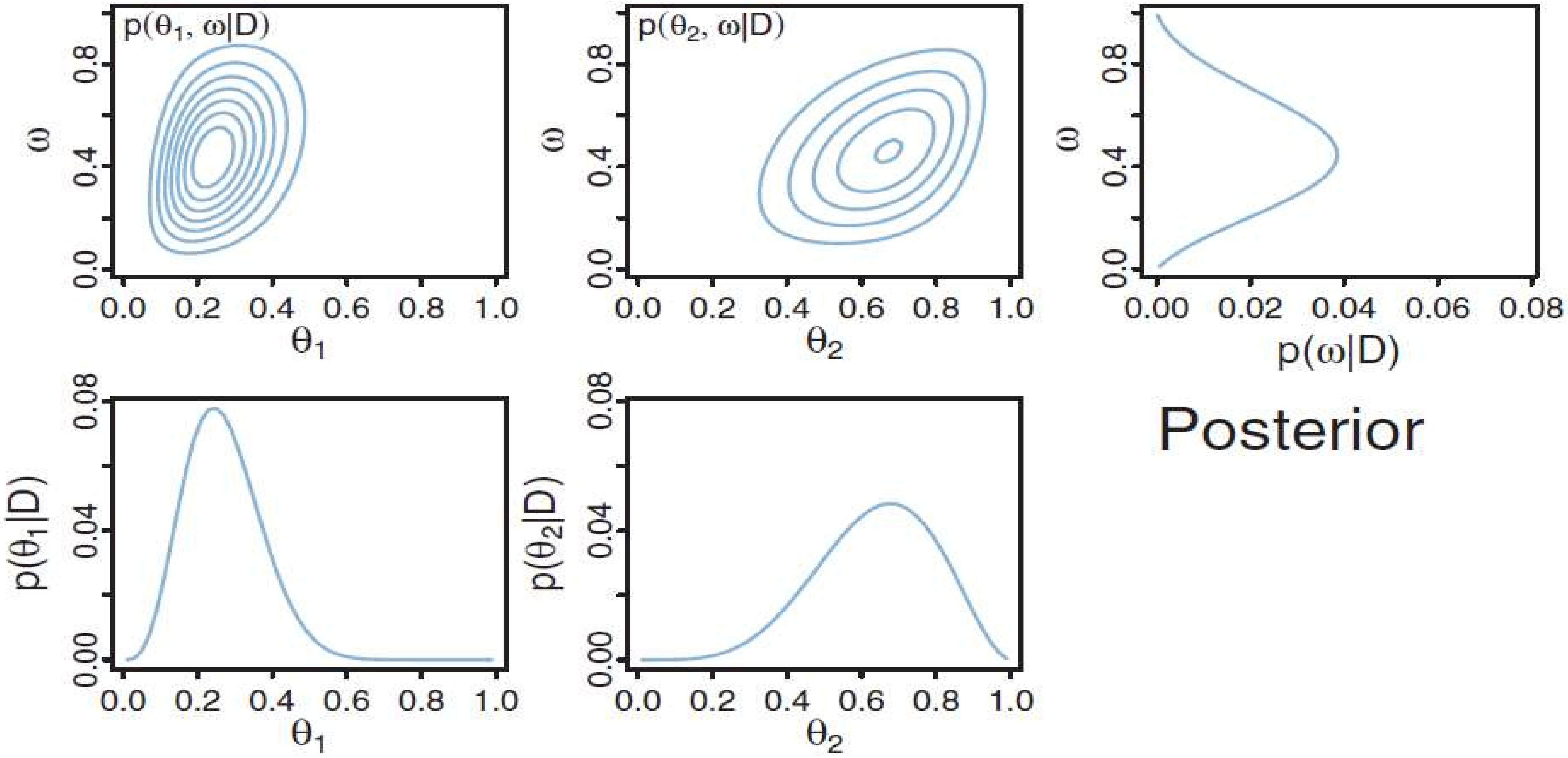
9.2. MULTIPLE COINS FROM A SINGLE MINT:

$p(\{y_{i|s}\}|\theta_s, \omega)$=$\mathrm{p}(\{\mathrm{y}\}|\theta_s)$→ likelihood function depends only on $\theta_s$and not on ω.



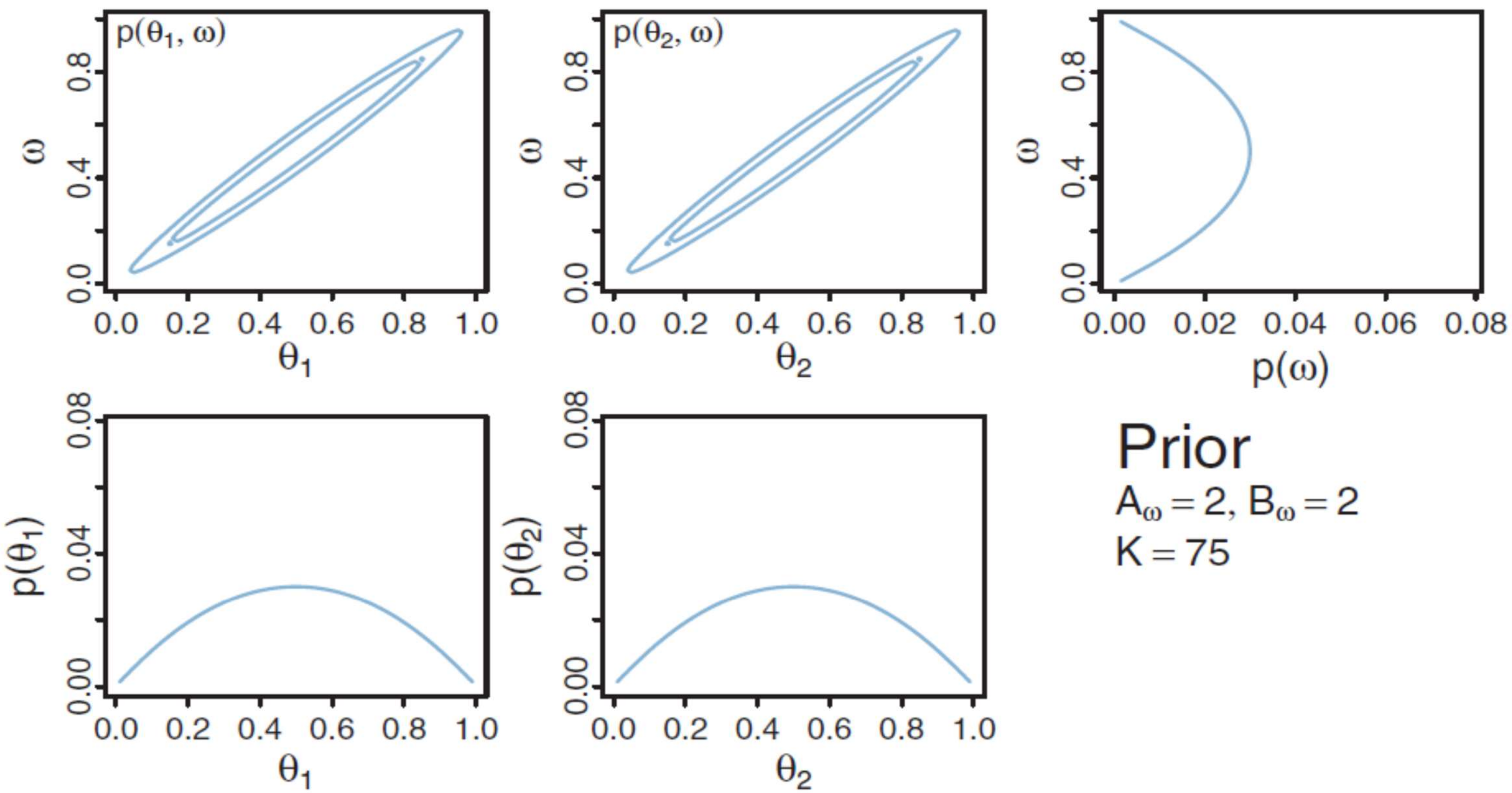
9.2. MULTIPLE CONIS FROM A SINGLE MINT:

- *p*(θ_1,*ω*|*D*) is centered near the proportion 3/15 = 0.2.
- *p*(θ_2,*ω*|*D*) is centered near the proportion 4/5 = 0.8.
- *p*(θ_1,*ω*|*D*) has less uncertainty *p*(θ_2,*ω*|*D*) , indicated by the widths of the distributions. Because θ_1 has larger sample size.



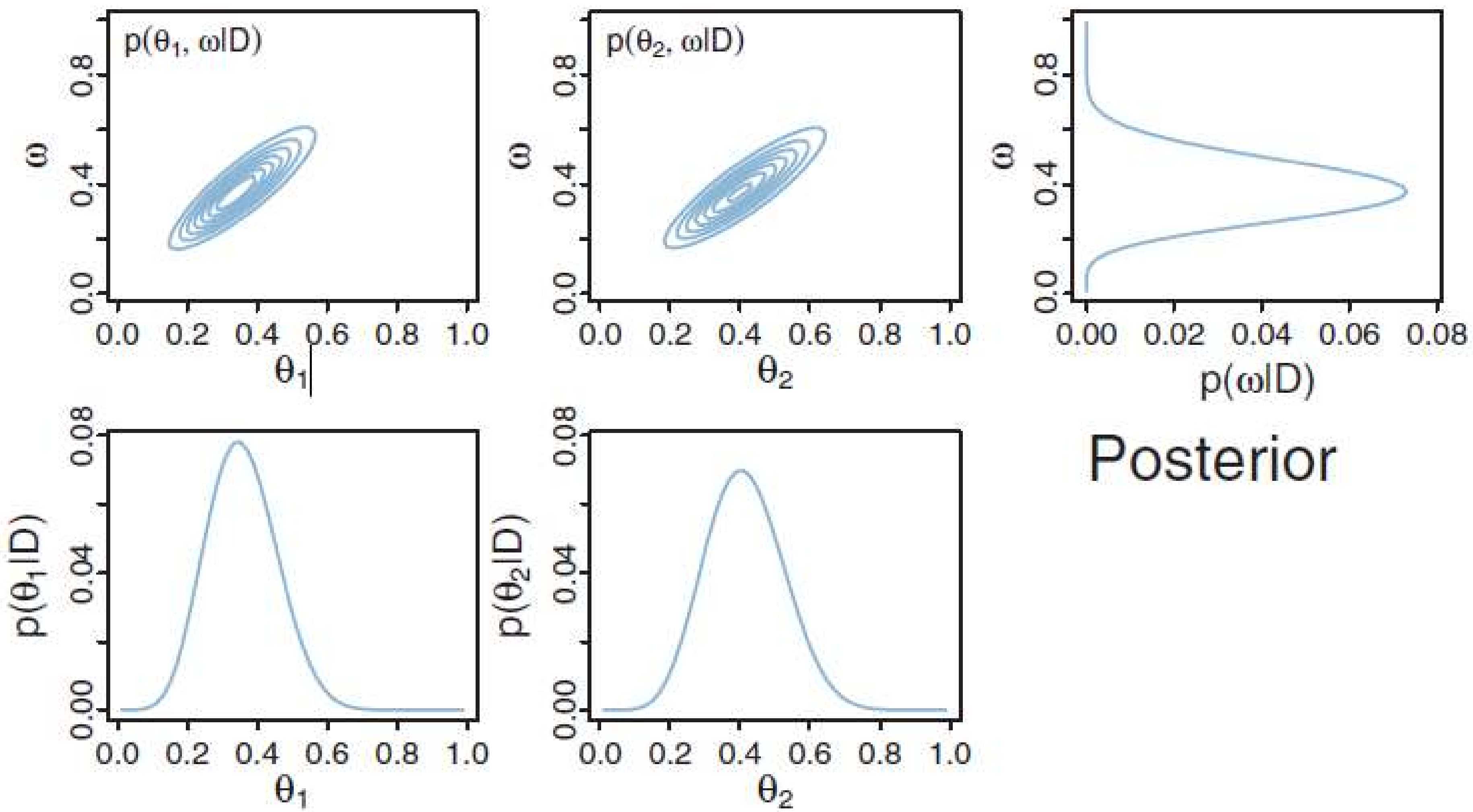
9.2. MULTIPLE CONIS FROM A SINGLE MINT:

- *K* =75 instead of K=5 →prior dependency of θ_s on ω is much stronger



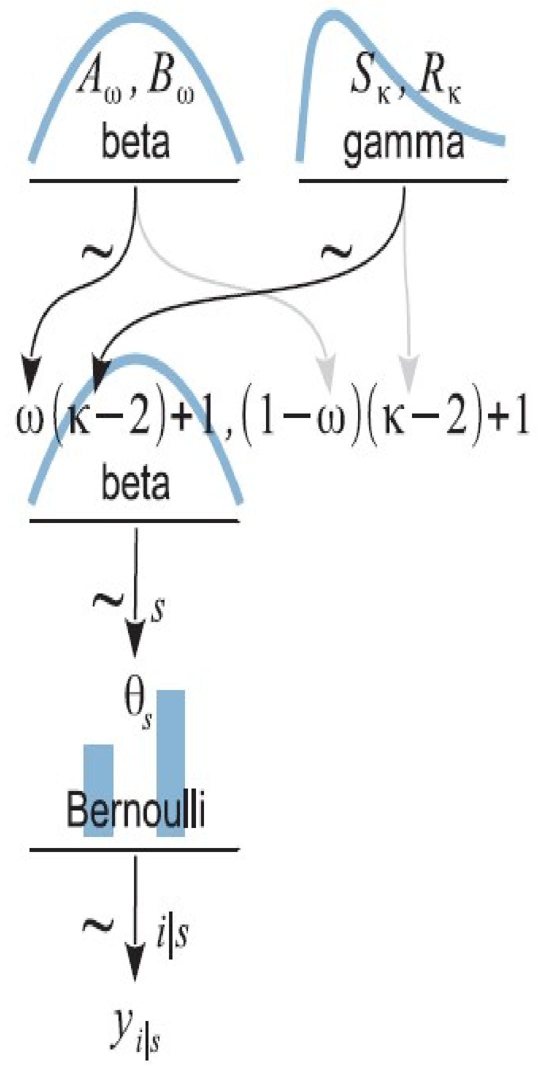
9.2. MULTIPLE CONIS FROM A SINGLE MINT:

- ω pulls all the estimates toward the focal zone.
- The posterior on θ_2 is peaked around 0.4, far from the proportion 4/5 = 0.8 in its coin's data!
- Less shift in θ_1 because it has larger sample size.



9.2.2. A realistic model with MCMC:

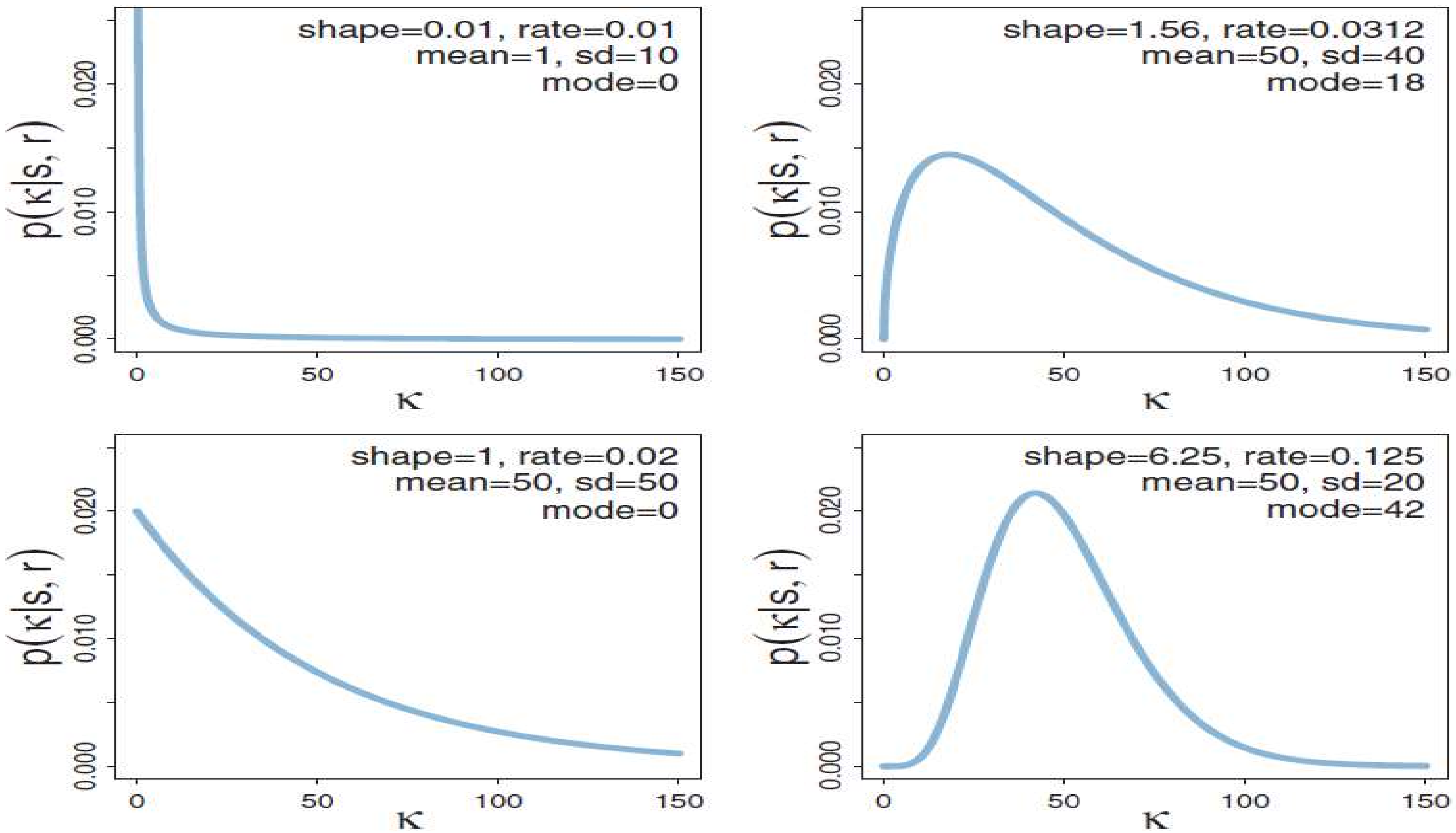
- In real situations, we do not know the value of *κ* in advance.
- Therefore, we expand the hierarchical model to include *κ*.
- If we want to estimate *κ*, we must have prior uncertainty about it.
- Because the value of *κ*−2 must be non-negative, the prior distribution on *κ*−2 must not allow negative values.
- gamma(*κ*|*s*, *r*) distribution is a probability density for *κ* ≥ 0.
- shape *s* and rate *r* parameters.



9.2.2. A realistic model with MCMC:

$$s = \frac{\mu^2}{\sigma^2} \quad \text{and} \quad r = \frac{\mu}{\sigma^2} \quad \text{for mean } \mu > 0$$

$$s = 1 + \omega r \quad \text{where} \quad r = \frac{\omega + \sqrt{\omega^2 + 4\sigma^2}}{2\sigma^2} \quad \text{for mode } \omega > 0$$



9.2.4. Example: Therapeutic touch:

- Rosa et al. (1998) investigated a key claim of practitioners of therapeutic touch, namely, that the practitioners can sense a body's energy field.
- If this is true, then practitioners should be able to sense **which of their hands is near another person's hand.** even without being able to see their hands.

Experiment:

a) The practitioner sat with her hands extended through cutouts in a cardboard screen, which prevented the practitioner from seeing.

b) the experimenter. On each trial, the experimenter flipped a coin and held her hand a few centimeters above one or the other of the practitioner's hands, as dictated by the flip of the coin.

c) The practitioner then responded with her best guess regarding which of her hand's was being hovered over.

d) Each trial was scored as correct or wrong.

9.2.4. Example: Therapeutic touch:

Experiment:

a) The practitioner sat with her hands extended through cutouts in a cardboard screen, which prevented the practitioner from seeing.

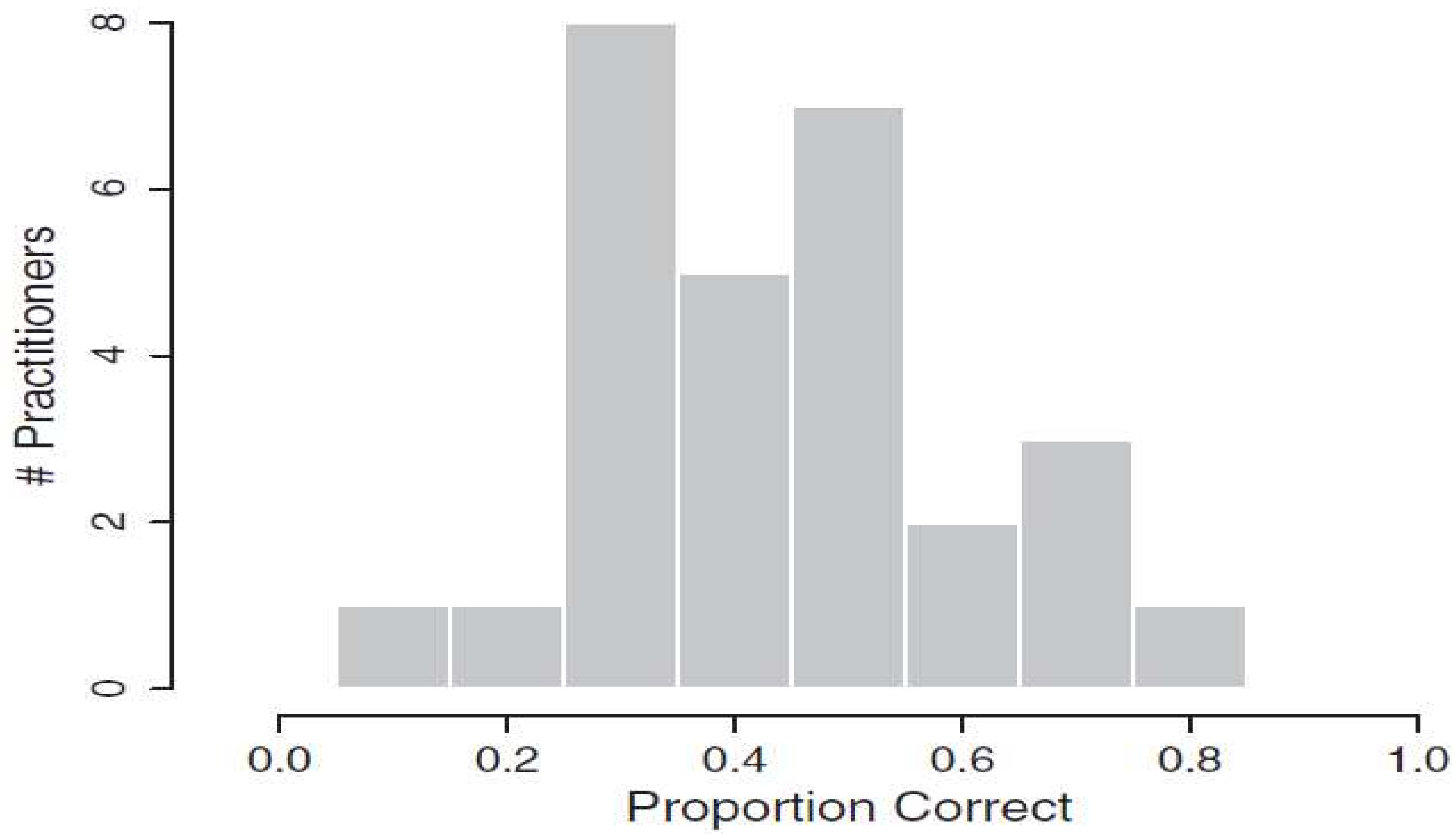
b) the experimenter. On each trial, the experimenter flipped a coin and held her hand a few centimeters above one or the other of the practitioner’s hands, as dictated by the flip of the coin.

c) The practitioner then responded with her best guess regarding which of her hand’s was being hovered over.

a) Each trial was scored as correct or wrong.

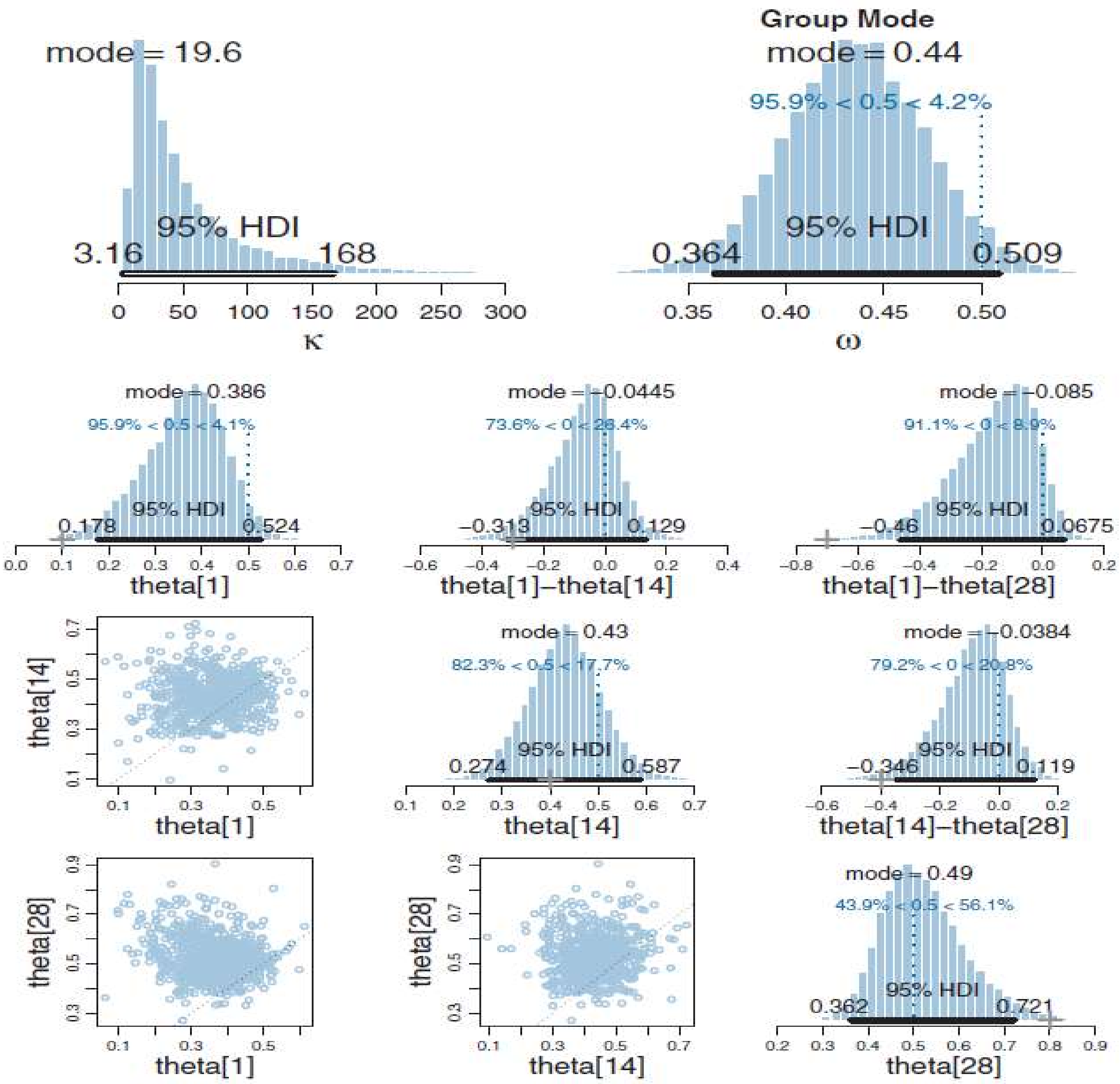
9.2.4. Example: Therapeutic touch:

- The proportions correct for the 28 subjects.
- Chance performance is 0.50.
- The question is how much the group as a whole differed from chance performance, and how much any individuals differed from chance performance.



9.2.4. Example: Therapeutic touch:

- The group-level mode, ω. It's most credible value is less than 0.5, and its 95% HDI includes the chance value of 0.5.

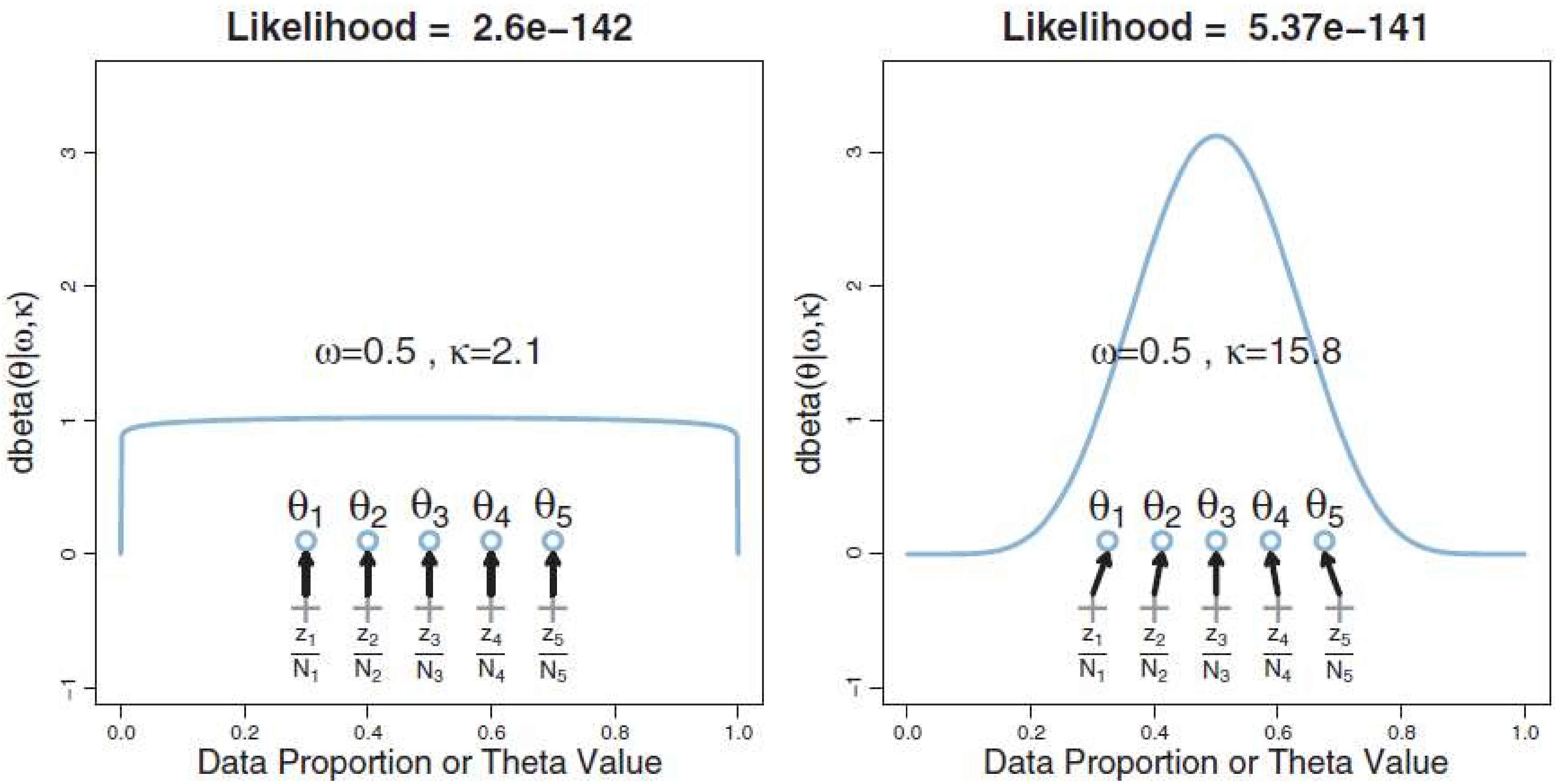


- The group-level parameters pull the estimate of the individual toward what is typical for the group.

.

9.3. SHRINKAGE IN HIERARCHICAL MODELS

- In typical hierarchical models, the estimates of low-level parameters are pulled closer together than they would be if there were not a higher-level distribution.
- This pulling together is called *shrinkage* of the estimates.



9.5. EXTENDING THE HIERARCHY: SUBJECTS WITHIN CATEGORIES:

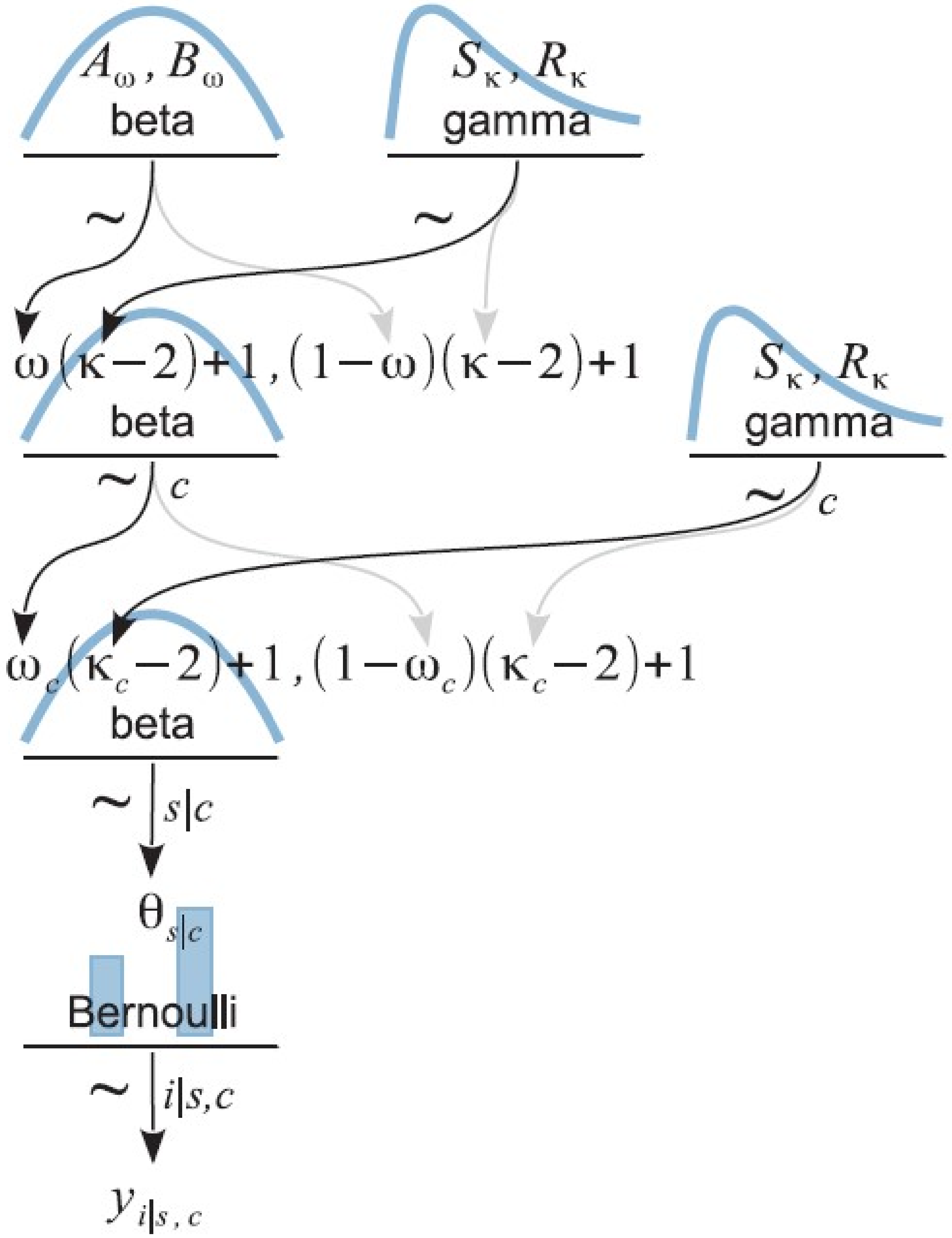
- Many data structures invite hierarchical descriptions that may have multiple levels.

Example: Consider professional baseball players who have different fielding positions, so it is meaningful to categorize players by their primary position.

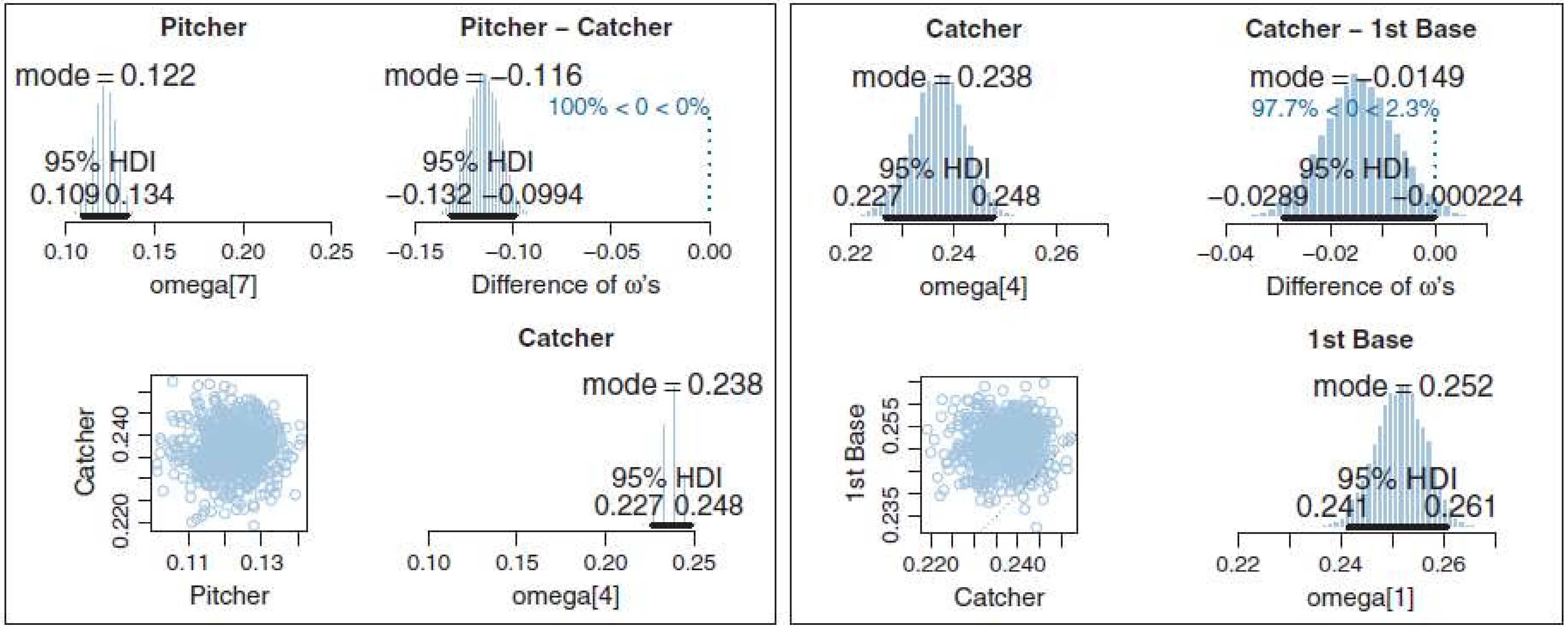
- We want estimate batting abilities for **individual players**, and for **positions**, and for the overarching **group of professional players.**

9.5. EXTENDING THE HIERARCHY: SUBJECTS WITHIN CATEGORIES:

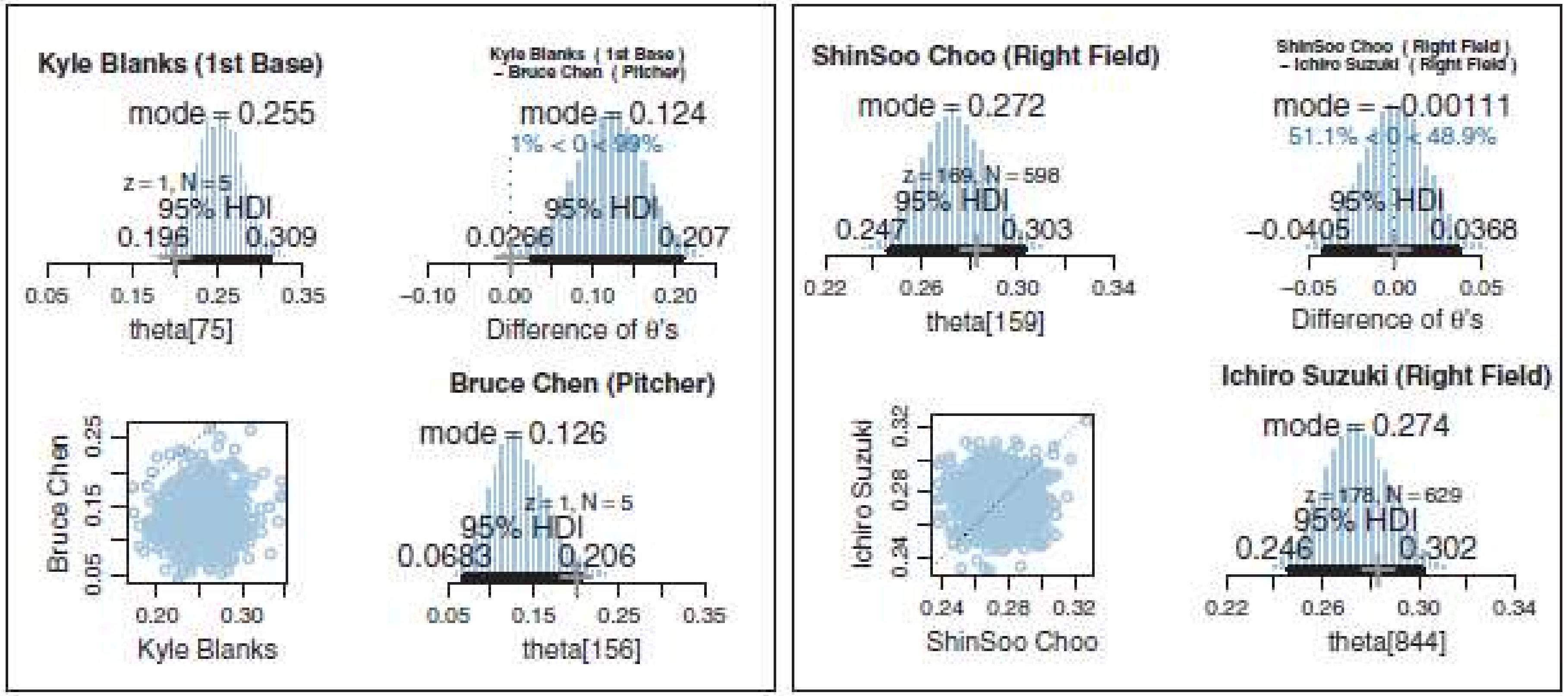
- Extra layer added for the category level.
- The model assumes that all the category modes come from a higher-level beta distribution that describes the variation across categories.
- We are estimating ω and κ, we must specify prior distributions for them, indicated at the



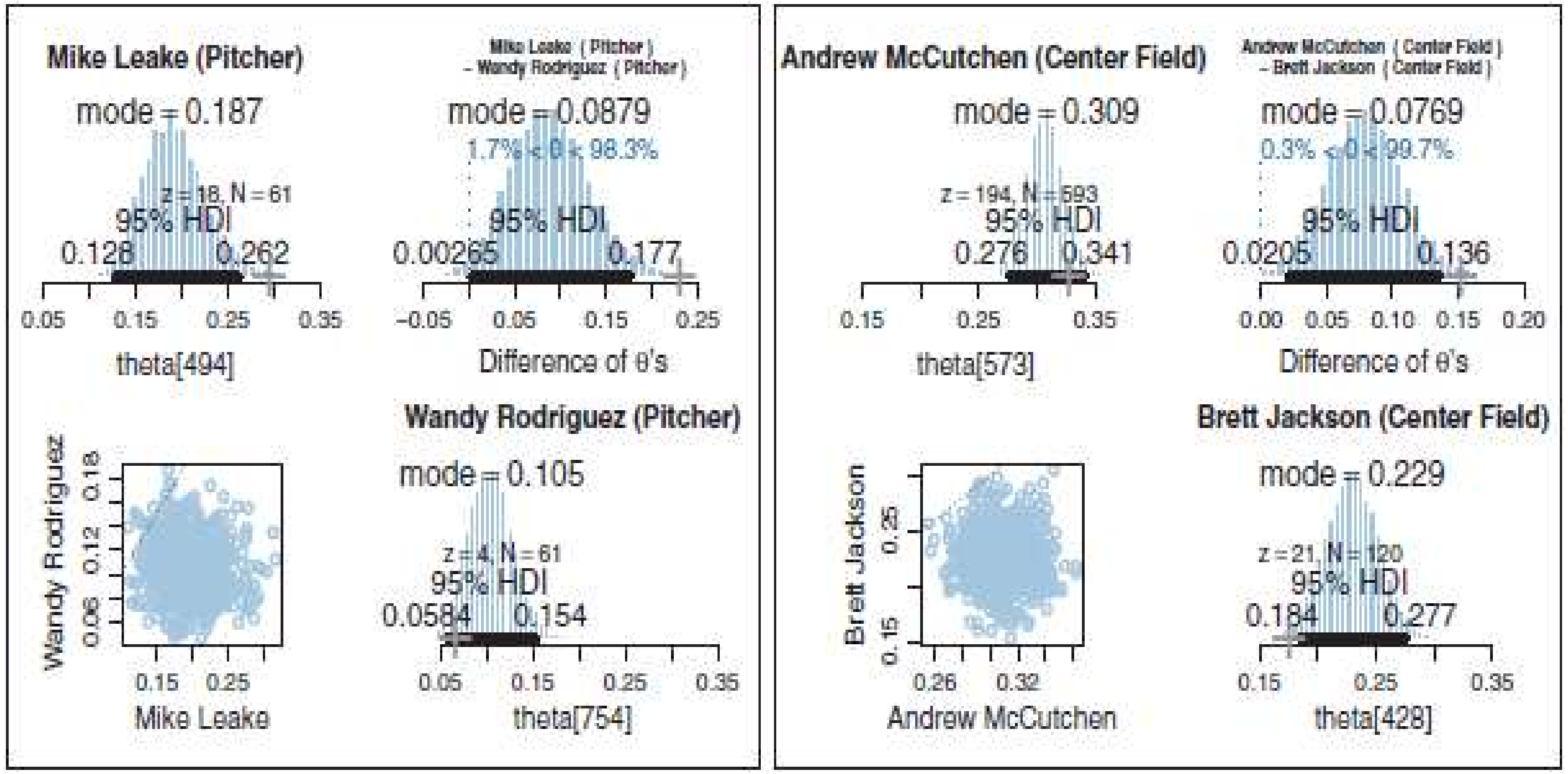
9.5.1. Example: Baseball batting abilities by position



9.5.1. Example: Baseball batting abilities by position



9.5.1. Example: Baseball batting abilities by position



Thank You