

Figure 6.8 Redisplay of Figure 6.7 without ordering the rows, columns, and persons in order of increasing response. Once again, the left column shows the observed data and the right columns show replicated datasets from the model. Without the ordering, it is difficult to notice the discrepancies between data and model, which are easily apparent in Figure 6.7.

6.8, which presents the same information as in Figure 6.7 but without the ordering. Here, the discrepancies between data and model are not clear at all.

Displaying summary statistics or inferences

A key principle of exploratory data analysis is to exploit regular structure to display data more effectively. The analogy in modeling is hierarchical or multilevel modeling, in which batches of parameters capture variation at different levels. When checking model fit, hierarchical structure can allow us to compare batches of parameters to their reference distribution. In this scenario, the replications correspond to new draws of a batch of parameters.

We illustrate with inference from a hierarchical model from psychology. This was a fairly elaborate model, whose details we do not describe here; all we need to know for this example is that the model included two vectors of parameters, ϕ_1, \dots, ϕ_{90} , and ψ_1, \dots, ψ_{69} , corresponding to patients and psychological symptoms, and that each of these 159 parameters were assigned independent Beta(2, 2) prior distributions. Each of these parameters represented a probability that a given patient or symptom is associated with a particular psychological syndrome.

Data were collected (measurements of which symptoms appeared in which patients) and the full Bayesian model was fitted, yielding posterior simulations for all these parameters. If the model were true, we would expect any single simulation draw of the vectors of patient parameters ϕ and symptom parameters ψ to look like independent draws from the Beta(2, 2) distribution. We know this because of the following reasoning:

- If the model were indeed true, we could think of the observed data vector y and the vector θ of the true values of all the parameters (including ϕ and ψ) as a random draw from their joint distribution, $p(y, \theta)$. Thus, y comes from the marginal distribution, the prior predictive distribution, $p(y)$.
- A single draw θ^s from the posterior inference comes from $p(\theta^s|y)$. Since $y \sim p(y)$, this

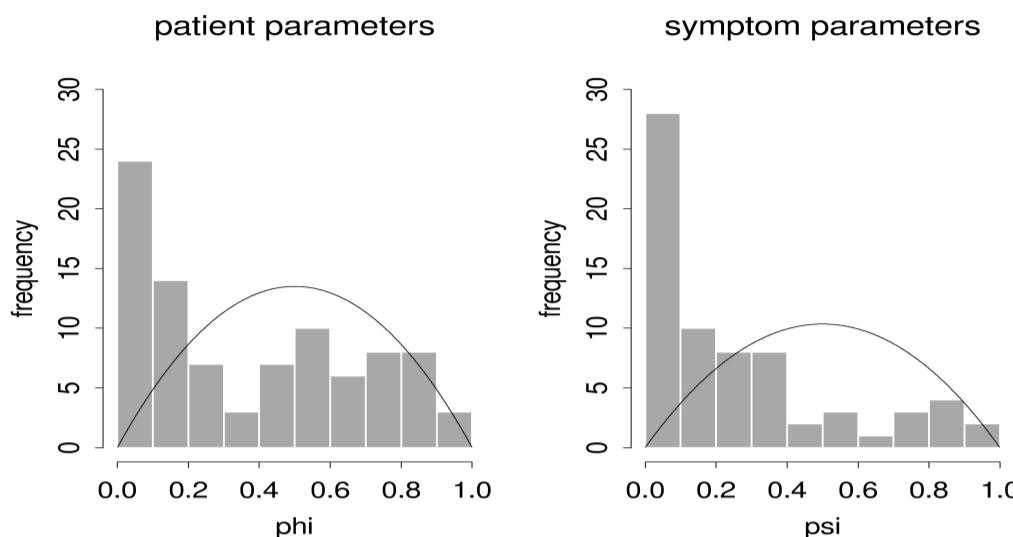


Figure 6.9 *Histograms of (a) 90 patient parameters and (b) 69 symptom parameters, from a single draw from the posterior distribution of a psychometric model. These histograms of posterior estimates contradict the assumed Beta(2,2) prior densities (overlain on the histograms) for each batch of parameters, and motivated us to switch to mixture prior distributions. This implicit comparison to the values under the prior distribution can be viewed as a posterior predictive check in which a new set of patients and a new set of symptoms are simulated.*

means that y, θ^s come from the model's joint distribution of y, θ , and so the marginal distribution of θ^s is the same as that of θ .

- That is, y, θ, θ^s have a combined joint distribution in which θ and θ^s have the same marginal distributions (and the same joint distributions with y).

Thus, as a model check we can plot a histogram of a single simulation of the vector of parameters ϕ or ψ and compare to the prior distribution. This corresponds to a posterior predictive check in which the inference from the observed data is compared to what would be expected if the model were applied to a new set of patients and a new set of symptoms.

Figure 6.9 shows histograms of a single simulation draw for each of ϕ and ψ as fitted to our dataset. The lines show the Beta(2,2) prior distribution, which clearly does not fit. For both ϕ and ψ , there are too many cases near zero, corresponding to patients and symptoms that almost certainly are not associated with a particular syndrome.

Our next step was to replace the offending Beta(2,2) prior distributions by mixtures of two beta distributions—one distribution with a spike near zero, and another that is uniform between 0 and 1—with different models for the ϕ 's and the ψ 's. The exact model is,

$$\begin{aligned} p(\phi_j) &= 0.5 \text{Beta}(\phi_j|1, 6) + 0.5 \text{Beta}(\phi_j|1, 1) \\ p(\psi_j) &= 0.5 \text{Beta}(\psi_j|1, 16) + 0.5 \text{Beta}(\psi_j|1, 1). \end{aligned}$$

We set the parameters of the mixture distributions to fixed values based on our understanding of the model. It was reasonable for these data to suppose that any given symptom appeared only about half the time; however, labeling of the symptoms is subjective, so we used beta distributions peaked near zero but with some probability of taking small positive values. We assigned the Beta(1,1) (that is, uniform) distributions for the patient and symptom parameters that were not near zero—given the estimates in Figure 6.9, these seemed to fit the data better than the original Beta(2,2) models. (The original reason for using Beta(2,2) rather than uniform prior distributions was so that maximum likelihood estimates would be in the interior of the interval [0, 1], a concern that disappeared when we moved to Bayesian inference; see Exercise 4.9.)

Some might object to revising the prior distribution based on the fit of the model to the data. It is, however, consistent with common statistical practice, in which a model is iteratively altered to provide a better fit to data. The natural next step would be to add