As an example, use the hot dog calorie data from Example 11.6.2. In this example, $p = 4$. We shall use a prior distribution in which $\lambda_0 = \alpha_0 = 1$, $\beta_0 = 0.1$, $u_0 = 0.001$, and $\psi_0 = 170$. We use $k = 6$ Markov chains and do $m = 100$ burn-in simulations, which turn out to be more than enough to make the maximum of all nine F statistics less than $1 + 0.44m$. We then run each of the six Markov chains another 10,000 iterations. The samples from the posterior distribution allow us to answer any questions that we might have about the parameters, including some that we would not have been able to answer using the analysis done in Chapter 11. For example, the posterior means and standard deviations of some of the parameters are listed in Table 12.6. To see how different the variances are, we can estimate the probability that the variance of one group is at least 2.25 times as high as that of another group by computing the fraction of iterations ℓ in which at least one $\tau_i^{(\ell)}/\tau_j^{(\ell)} > 2.25$. The result is 0.4, indicating that there is some chance that at least some of the variances are different. If the variances are different, the ANOVA calculations in Chapter 11 are not justified.

We can also address the question of how much difference there is between the μ_i's. For comparison, we shall do the same calculations that we did in Example 12.3.7. In 99 percent of the 60,000 simulations, at least one $|\mu_i^{(\ell)} - \mu_j^{(\ell)}| > 26.35$. In about one-half of the simulations, all $|\mu_i^{(\ell)} - \mu_j^{(\ell)}| > 2.224$. And in 99 percent of the simulations, the average of the differences was at least 13.78. Figure 12.9 contains a plot of the sample c.d.f.'s of the largest, smallest, and average of the six $|\mu_i - \mu_j|$ differences. Careful examination of the results in this example shows that the four μ_i's appear to be closer together than we would have thought after the analysis of Example 12.3.7. This is typical of what occurs when we use a proper prior in a hierarchical model. In Example 12.3.7, the μ_i's were all independent, and they did not have a common unknown mean in the prior. In Example 12.5.6, the μ_i's all have a common prior distribution with mean ψ, which is an additional unknown parameter. The estimation of this additional parameter allows the posterior distributions of the μ_i's to be pulled toward a

Table 12.6 Posterior means and standard deviations for some parameters in Example 12.5.6

Type	***Beef***	***Meat***	***Poultry***	***Specialty***
i	1	2	3	4
$E(\mu_i\|\boldsymbol{y})$	156.6	158.3	120.5	159.6
$(Var(\mu_i\|\boldsymbol{y}))^{1/2}$	4.893	5.825	5.521	7.615
$E(1/\tau_i\|\boldsymbol{y})$	495.6	608.5	542.9	568.2
$(Var(1/\tau_i\|\boldsymbol{y}))^{1/2}$	166.0	221.2	201.6	307.4
$E(\psi\|\boldsymbol{y}) = 151.0$		$(Var(\psi\|\boldsymbol{y}))^{1/2} = 11.16$		

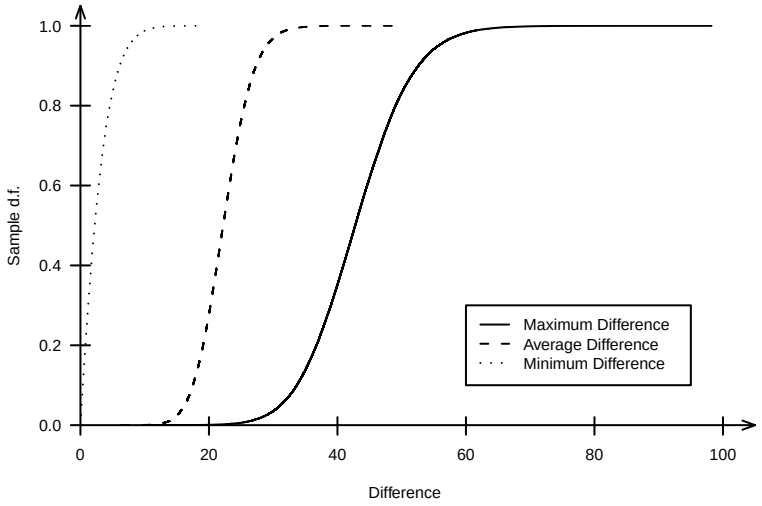

Figure 12.9 Sample c.d.f.'s of the maximum, average, and minimum of the six $|\mu_i - \mu_j|$ differences for Example 12.5.6.

location that is near the average of all of the samples. With these data, the overall sample average is 147.60. ◀

Prediction

All of the calculations done in the examples of this section have concerned functions of the parameters. The sample from the posterior distribution that we obtain from Gibbs sampling can also be used to make predictions and form prediction intervals for future observations. The most straightforward way to make predictions is to simulate the future data conditional on each value of the parameter from the posterior sample. Although there are more efficient methods for predicting, this method is easy to describe and evaluate.

Example 12.5.7 **Calories in Hot Dogs.** In Example 12.5.6, we might be concerned with how different we should expect the calorie counts of two hot dogs to be. For example, let Y_1 and Y_3 be future calorie counts for hot dogs of the beef and poultry varieties, respectively. We can form a prediction interval for $D = Y_1 - Y_3$ as follows. For each iteration ℓ, let the simulated parameter vector be

$$\theta^{(\ell)} = \left(\mu_1^{(\ell)}, \mu_2^{(\ell)}, \mu_3^{(\ell)}, \mu_4^{(\ell)}, \tau_1^{(\ell)}, \tau_2^{(\ell)}, \tau_3^{(\ell)}, \tau_4^{(\ell)}, \psi^{(\ell)}, \beta^{(\ell)}\right).$$

For each ℓ, simulate a beef hot dog calorie count $Y_1^{(\ell)}$ having the normal distribution with mean $\mu_1^{(\ell)}$ and variance $1/\tau_1^{(\ell)}$. Also simulate a poultry hot dog calorie count $Y_3^{(\ell)}$ having the normal distribution with mean $\mu_3^{(\ell)}$ and variance $1/\tau_3^{(\ell)}$. Then compute $D^{(\ell)} = Y_1^{(\ell)} - Y_3^{(\ell)}$. Sample quantiles of the values $D^{(1)}, \ldots, D^{(60000)}$ can be used to estimate quantiles of the distribution of D.

For example, suppose that we want a 90 percent prediction interval for D. We simulate 60,000 $D^{(\ell)}$ values as above and find the 0.05 and 0.95 sample quantiles to be -18.49 and 90.63, which are then the endpoints of our prediction interval. To assess how close the simulation estimators are to the actual quantiles of the distribution of D, we compute the simulation standard errors of the two endpoints. For the samples from each of the $k = 6$ Markov chains, we can compute the sample 0.05 quantiles of our D values. We can then use these values as $Z_1, \ldots, Z_6$ in Eq. (12.5.1) to compute a value S. Our simulation standard error is then $S/6^{1/2}$. We can then repeat this for the sample 0.95 quantiles. For the two endpoints of our interval, the simulation standard errors are 0.2228 and 0.4346, respectively. These simulation standard errors are fairly small compared to the length of the prediction interval. ◀