of data, each of which can be modeled using a DP, and we begin by describing the HDP at this level of generality, subsequently specializing to the HMM.

To describe the HDP, suppose there are J groups of data and let $\{y_{j1}, \dots, y_{jN_j}\}$ denote the set of observations in group j. Assume that there are a collection of DP mixture models underlying the observations in these groups:

$$\begin{aligned} G_j = \sum_{t=1}^{\infty} \tilde{\pi}_{jt} \delta_{\theta^*_{jt}}, \qquad & \tilde{\pi}_j | \alpha \sim \mathrm{GEM}(\alpha), j = 1, \dots, J, \\ & \theta^*_{jt} | G_0, \sim G_0, t = 1, 2, \dots, \\ \theta'_{ji} | G_j \sim G_j, \qquad y_{ji} | \theta'_{ji} \sim F(\theta'_{ji}), & \qquad j = 1, \dots, J, i = 1, \dots, N_j. \end{aligned} \tag{4.1}$$

We wish to tie the DP mixtures across the different groups such that atoms that underly the data in group j can be used in group j'. The problem is that if G_0 is absolutely continuous with respect to the Lebesgue measure (as it generally is for continuous parameters), then the atoms in G_j will be distinct from those in $G_{j'}$ with probability one. The solution to this problem is to let G_0 itself be a draw from a DP:

$$\begin{aligned} G_0 = \sum_{k=1}^{\infty} \beta_k \delta_{\theta_k}, \qquad & \beta | \gamma \sim \mathrm{GEM}(\gamma), \\ & \theta_k | H, \lambda \sim H(\lambda), k = 1, 2, \dots. \end{aligned} \tag{4.2}$$

In this hierarchical model, G_0 is atomic and random. Letting G_0 be a base measure for the draw $G_j \sim \mathrm{DP}(\alpha, G_0)$ implies that only these atoms can appear in G_j. Thus, atoms can be shared among the collection of random measures $\{G_j\}$. The HDP model is depicted graphically in two different ways in Figure 4(c) and (d).

Teh et al. (2006) have also described the marginal probabilities obtained from integrating over the random measures G_0 and $\{G_j\}$. They show that these marginals can be described in terms of a *Chinese restaurant franchise* (CRF) that is an analog of the Chinese restaurant process. The CRF is comprised of J restaurants, each corresponding to an HDP group, and an infinite buffet line of dishes common to all restaurants. The process of seating customers at tables, however, is restaurant specific. Each customer is preassigned to a given restaurant determined by that customer's group j. Upon entering the jth restaurant in the CRF, customer y_{ji} sits at currently occupied tables t_{ji} with probability proportional to the number of currently seated customers, or starts a new table $T_j + 1$ with probability proportional to α. The first customer to sit at a table goes to the buffet line and picks a dish k_{jt} for their table, choosing the dish with probability proportional to the number of times that dish has been picked previously, or ordering a new dish θ_{K+1} with probability proportional to γ. The intuition behind this predictive distribution is that integrating over the global dish probabilities β results in customers making decisions based on the observed popularity of the dishes throughout the entire franchise. See the Supplementary Material for further details [Fox et al. (2010)].