denoted by $\beta \sim \mathrm{GEM}(\gamma)$. With probability one, a random draw $G_0 \sim \mathrm{DP}(\gamma, H)$ can be expressed as

$$G_0 = \sum_{k=1}^{\infty} \beta_k \delta_{\theta_k}, \qquad \theta_k | H \sim H, k = 1, 2, \dots, \tag{3.3}$$

where δ_θ denotes a unit-mass measure concentrated at θ and where $\{\theta_k\}$ are drawn independently from H. From this definition, we see that the DP actually defines a distribution over discrete probability measures. The stick-breaking construction also gives us insight into how the concentration parameter γ controls the relative magnitude of the mixture weights β_k, and thus determines the model complexity in terms of the expected number of components with significant probability mass.

The DP has a number of properties which make inference based on this nonparametric prior computationally tractable. Consider a set of observations $\{\theta'_i\}$ with $\theta'_i \sim G_0$. Because probability measures drawn from a DP are discrete, there is a strictly positive probability of multiple observations θ'_i taking identical values within the set $\{\theta_k\}$, with θ_k defined as in equation (3.3). For each value θ'_i, let z_i be an indicator random variable that picks out the unique value k such that $\theta'_i = \theta_{z_i}$. Blackwell and MacQueen (1973) introduced a Pólya urn representation of the θ'_i:

$$\begin{aligned} \theta'_i | \theta'_1, \dots, \theta'_{i-1} &\sim \frac{\gamma}{\gamma + i - 1} H + \sum_{j=1}^{i-1} \frac{1}{\gamma + i - 1} \delta_{\theta'_j} \\ &= \frac{\gamma}{\gamma + i - 1} H + \sum_{k=1}^{K} \frac{n_k}{\gamma + i - 1} \delta_{\theta_k}, \end{aligned} \tag{3.4}$$

implying the following predictive distribution for the indicator random variables:

$$p(z_{N+1} = z | z_1, \dots, z_N, \gamma) = \frac{\gamma}{N + \gamma} \delta(z, K+1) + \frac{1}{N + \gamma} \sum_{k=1}^{K} n_k \delta(z, k). \tag{3.5}$$

Here, $n_k = \sum_{i=1}^{N} \delta(z_i, k)$ is the number of indicator random variables taking the value k, and $K + 1$ is a previously unseen value. We use the notation $\delta(z, k)$ to indicate the discrete Kronecker delta. This representation can be used to sample observations from a DP without explicitly constructing the countably infinite random probability measure $G_0 \sim \mathrm{DP}(\gamma, H)$.

The distribution on partitions induced by the sequence of conditional distributions in equation (3.5) is commonly referred to as the *Chinese restaurant process*. The analogy, which is useful in developing various generalizations of the Dirichlet process we consider in this paper, is as follows. Take i to be a customer entering a restaurant with infinitely many tables, each serving a unique dish θ_k. Each arriving customer chooses a table, indicated by z_i, in proportion to how many customers are currently sitting at that table. With some positive probability proportional to γ, the customer starts a new, previously unoccupied table $K + 1$. The Chinese