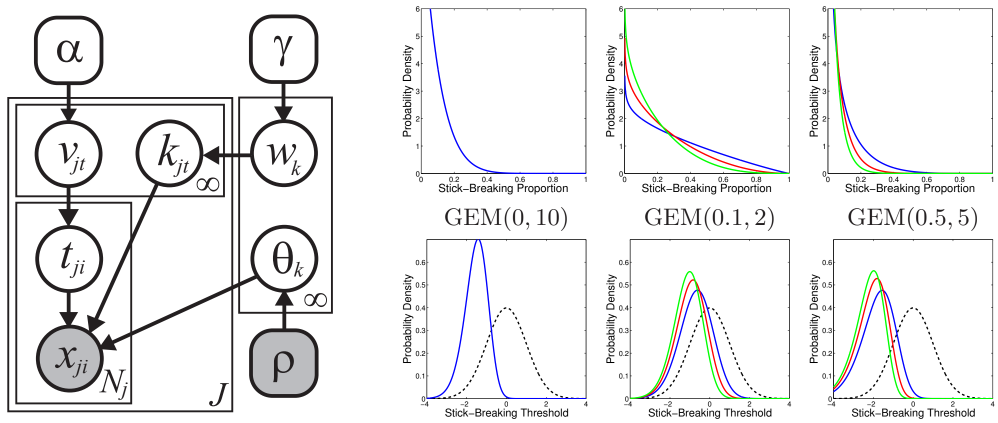

Figure 2: Stick-breaking representation of a hierarchical Pitman–Yor (HPY) model for J groups of features. *Left:* Directed graphical model in which global category frequencies $\boldsymbol{\varphi} \sim \mathrm{GEM}(\gamma)$ are constructed from stick-breaking proportions $w_k \sim \mathrm{Beta}(1-\gamma_a, \gamma_b + k\gamma_a)$, as in Eq. (1). Similarly, $v_{jt} \sim \mathrm{Beta}(1-\alpha_a, \alpha_b + t\alpha_a)$ define region areas $\boldsymbol{\pi}_j \sim \mathrm{GEM}(\alpha)$ for image j. Each of the N_j features x_{ji} is independently sampled as in Eq. (2). *Upper right:* Beta distributions from which stick proportions w_k are sampled for three different PY processes: $k=1$ (blue), $k=10$ (red), $k=20$ (green). *Lower right:* Corresponding distributions on thresholds for an equivalent generative model employing zero mean, unit variance Gaussians (dashed black). See Sec. 4.1.

3.2 Variational Learning for HPY Mixture Models

To allow efficient learning of HPY model parameters from large image databases, we have developed a mean field variational method which combines and extends previous approaches for DP mixtures [21, 22] and finite topic models. Using the stick-breaking representation of Fig. 2, and a factorized variational posterior, we optimize the following lower bound on the marginal likelihood:

$$\log p(\mathbf{x} \mid \alpha, \gamma, \rho) \geq H(q) + \mathbb{E}_q[\log p(\mathbf{x}, \mathbf{k}, \mathbf{t}, \mathbf{v}, \mathbf{w}, \boldsymbol{\theta} \mid \alpha, \gamma, \rho)] \tag{3}$$

$$q(\mathbf{k}, \mathbf{t}, \mathbf{v}, \mathbf{w}, \boldsymbol{\theta}) = \left[\prod_{k=1}^{K} q(w_k \mid \omega_k) q(\theta_k \mid \eta_k)\right] \cdot \prod_{j=1}^{J} \left[\prod_{t=1}^{T} q(v_{jt} \mid \nu_{jt}) q(k_{jt} \mid \kappa_{jt})\right] \prod_{i=1}^{N_j} q(t_{ji} \mid \tau_{ji})$$

Here, $H(q)$ is the entropy. We *truncate* the variational posterior [21] by setting $q(v_{jT}=1)=1$ for each image or group, and $q(w_K = 1) = 1$ for the shared global clusters. Multinomial assignments $q(k_{jt} \mid \kappa_{jt})$, $q(t_{ji} \mid \tau_{ji})$, and beta stick proportions $q(w_k \mid \omega_k)$, $q(v_{jt} \mid \nu_{jt})$, then have closed form update equations. To avoid bias, we sort the current sets of image segments, and global categories, in order of decreasing aggregate assignment probability after each iteration [22].

4 Segmentation with Spatially Dependent Pitman–Yor Processes

We now generalize the HPY image segmentation model of Fig. 2 to capture spatial dependencies. For simplicity, we consider a single-image model in which features x_i are assigned to regions by indicator variables z_i, and each segment k has its own appearance parameters θ_k (see Fig. 3). As in Sec. 3.1, however, this model is easily extended to share appearance parameters among images.

4.1 Coupling Assignments using Thresholded Gaussian Processes

Consider a generative model which partitions data into two clusters via assignments $z_i \in \{0, 1\}$ sampled such that $\mathbb{P}[z_i = 1] = v$. One representation of this sampling process first generates a Gaussian auxiliary variable $u_i \sim \mathcal{N}(0,1)$, and then chooses z_i according to the following rule:

$$z_i = \begin{cases} 1 & \text{if } u_i < \Phi^{-1}(v) \\ 0 & \text{otherwise} \end{cases} \qquad\qquad \Phi(u) \triangleq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-s^2/2}\, ds \tag{4}$$

Here, $\Phi(u)$ is the standard normal *cumulative distribution function* (CDF). Since $\Phi(u_i)$ is uniformly distributed on $[0,1]$, we immediately have $\mathbb{P}[z_i = 1] = \mathbb{P}\big[u_i < \Phi^{-1}(v)\big] = \mathbb{P}[\Phi(u_i) < v] = v$.

We adapt this idea to PY processes using the stick-breaking representation of Eq. (1). In particular, we note that if $z_i \sim \boldsymbol{\pi}$ where $\pi_k = v_k \prod_{\ell=1}^{k-1}(1 - v_\ell)$, a simple induction argument shows that $v_k = \mathbb{P}[z_i = k \mid z_i \neq k-1, \ldots, 1]$. The stick-breaking proportion v_k is thus the *conditional* probability of choosing cluster k, given that clusters with indexes $\ell < k$ have been rejected. Combining