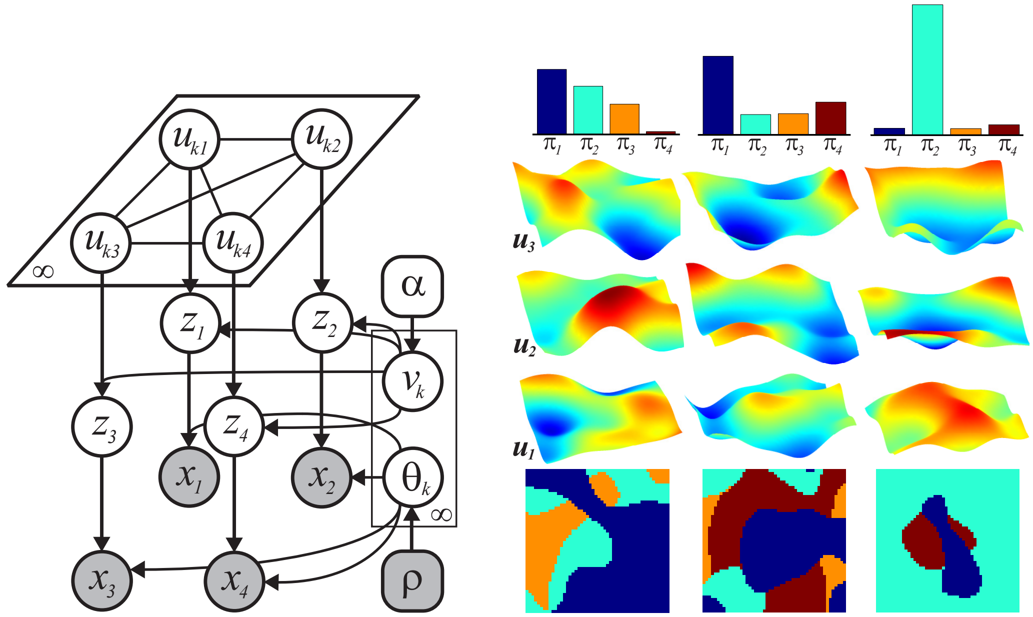

Figure 3: A nonparametric Bayesian approach to image segmentation in which thresholded Gaussian processes generate spatially dependent Pitman–Yor processes. *Left:* Directed graphical model in which expected segment areas $\boldsymbol{\pi} \sim \mathrm{GEM}(\alpha)$ are constructed from stick-breaking proportions $v_k \sim \mathrm{Beta}(1-\alpha_a, \alpha_b + k\alpha_a)$. Zero mean Gaussian processes ($u_{ki} \sim \mathcal{N}(0,1)$) are cut by thresholds $\Phi^{-1}(v_k)$ to produce segment assignments z_i, and thereby features x_i. *Right:* Three randomly sampled image partitions (columns), where assignments (bottom, color-coded) are determined by the *first* of the ordered Gaussian processes $\mathbf{u}_k$ to cross $\Phi^{-1}(v_k)$.

this insight with Eq. (4), we can generate samples $z_i \sim \boldsymbol{\pi}$ as follows:

$$z_i = \min\left\{k \mid u_{ki} < \Phi^{-1}(v_k)\right\} \qquad \text{where } \; u_{ki} \sim \mathcal{N}(0,1) \; \text{ and } \; u_{ki} \perp u_{\ell i}, k \neq \ell \tag{5}$$

As illustrated in Fig. 3, each cluster k is now associated with a zero mean *Gaussian process* (GP) $\mathbf{u}_k$, and assignments are determined by the sequence of *thresholds* in Eq. (5). If the GPs have identity covariance functions, we recover the basic HPY model of Sec. 3.1. More general covariances can be used to encode the prior probability that each feature pair occupies the same segment. Intuitively, the ordering of segments underlying this dependent PY model is analogous to *layered* appearance models [23], in which foreground layers *occlude* those that are farther from the camera.

To retain the power law prior on segment sizes justified in Sec. 2.3, we transform priors on stick proportions $v_k \sim \mathrm{Beta}(1-\alpha_a, \alpha_b + k\alpha_a)$ into corresponding random thresholds:

$$p(\bar{v}_k \mid \alpha) = \mathcal{N}(\bar{v}_k \mid 0,1) \cdot \mathrm{Beta}(\Phi(\bar{v}_k) \mid 1-\alpha_a, \alpha_b + k\alpha_a) \qquad\qquad \bar{v}_k \triangleq \Phi^{-1}(v_k) \tag{6}$$

Fig. 2 illustrates the threshold distributions corresponding to several different PY stick-breaking priors. As the number of features N becomes large relative to the GP covariance length-scale, the proportion assigned to segment k approaches π_k, where $\boldsymbol{\pi} \sim \mathrm{GEM}(\alpha_a, \alpha_b)$ as desired.

4.2 Variational Learning for Dependent PY Processes

Substantial innovations are required to extend the variational method of Sec. 3.2 to the Gaussian processes underlying our dependent PY processes. Complications arise due to the threshold assignment process of Eq. (5), which is "stronger" than the likelihoods typically used in probit models for GP classification, as well as the non-standard threshold prior of Eq. (6). In the simplest case, we place factorized Gaussian variational posteriors on thresholds $q(\bar{v}_k) = \mathcal{N}(\bar{v}_k \mid \nu_k, \delta_k)$ and assignment surfaces $q(u_{ki}) = \mathcal{N}(u_{ki} \mid \mu_{ki}, \lambda_{ki})$, and exploit the following key identities:

$$\mathbb{P}_q[u_{ki} < \bar{v}_k] = \Phi\left(\frac{\nu_k - \mu_{ki}}{\sqrt{\delta_k + \lambda_{ki}}}\right) \qquad \mathbb{E}_q[\log \Phi(\bar{v}_k)] \leq \log \mathbb{E}_q[\Phi(\bar{v}_k)] = \log \Phi\left(\frac{\nu_k}{\sqrt{1+\delta_k}}\right) \tag{7}$$

The first expression leads to closed form updates for Dirichlet appearance parameters $q(\theta_k \mid \eta_k)$, while the second evaluates the beta normalization constants in Eq. (6). We then *jointly* optimize each layer's threshold $q(\bar{v}_k)$ and assignment surface $q(\mathbf{u}_k)$, fixing all other layers, via backtracking conjugate gradient (CG) with line search. For details and further refinements, see [17].