3.3 Infinite Sparse Factor Models

We first briefly review Principal Components Analysis (PCA), Independent Components Analysis (ICA), and Factor Analysis (FA). Let $y_n \in \Re^D$ be one observed data point and $x_n \in \Re^K$ be a vector of hidden components (both of them are row vectors). Based on these models, we can explain y_n in terms of a linear superposition of x_n as

$$y_n = x_n \mathbf{G} + \epsilon_n, \tag{7}$$

where $\mathbf{G}$ is the factor loading matrix and ϵ_n is a noise vector. PCA and FA assume that the latent factors come from normal priors. The difference between the two models is that in PCA, the noise is assumed to be isotropic, while in FA, the noise is assumed to be diagonal. In ICA, the latent factors are assumed to be heavy-tailed,

Knowles and Ghahramani [13, 14] propose a Bayesian nonparametric extension of these models by introducing the IBP into the models. First, the IBP introduces sparsity in the models. Sparsity is desirable in many applications mainly because it improves predictive performance and makes the models more interpretable. Second, the IBP enables the models to automatically decide how many hidden components to use. We illustrate the graphical model of the model in Figure 3.

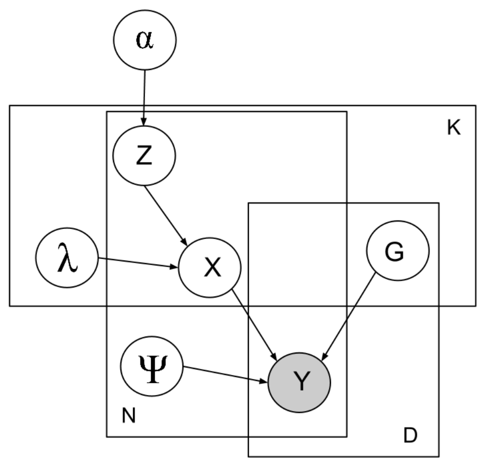

Figure 3: The graphical model of the infinite sparse factor analysis

Let $\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{E}$ be concatenated matrices of x_n, y_n, z_n and ϵ_n, respectively, and $\odot$ denote an element-wise multiplication. The model places the IBP prior on $\mathbf{Z}$ both to allow an infinite number of columns and to induce sparsity of the latent factors, $\mathbf{X}$. That is,

$$\mathbf{Y} = (\mathbf{X} \odot \mathbf{Z})\mathbf{G} + \mathbf{E}, \tag{8}$$

where

$$\begin{aligned} \epsilon_n &\sim N(0, \Psi_n) & \Psi_n &\sim IG(a, b) \\ Z &\sim IBP(\alpha, 1) & \alpha &\sim Gamma(e, f) \\ x_{nk} &\sim Z_{nk} N(0, \lambda_k^{-1}) + (1 - Z_{nk})\delta_0(x_{nk}) & \lambda_k &\sim Gamma(c, d) \\ g_{kd} &\sim N(0, 1), \end{aligned}$$

and δ_0 is a delta function at 0.

Recently, Paisley et al. [19] propose the beta process factor analysis. In this model, they use the beta-Bernoulli process approximation as a prior to Z:

$$\begin{aligned} z_{nk} &\sim Bernoulli(p_k) \\ p_k &\sim Beta(a/K, b(K-1)/K) \end{aligned}$$