A Stick-Breaking Likelihood for Categorical Data Analysis with Latent Gaussian Models
Mohammad Emtyayez Khan¹, Shakir Mohamed¹, Benjamin M. Marlin² and Kevin P. Murphy¹
¹University of British Columbia, ²University of Massachusetts, Amherst

Introduction
Motivation: Analysis of high-dimensional categorical data is essential in applications such as recommender systems, econometrics, social sciences, and medical diagnostics. Such analysis can be carried out using latent Gaussian models, which include multinomial log-regression, multi-class Gaussian process classification, categorial factor analysis, etc.

Problem: Parameter learning is difficult since the marginal likelihood contains an intractable integral, which arises due to the non-conjugacy between the likelihood and Gaussian prior on the latent variables.

Solution: We propose a novel stick-breaking likelihood for categorical data analysis and derive tractable and accurate lower bounds for the marginal likelihood. Our results demonstrate that the proposed stick-breaking model effectively captures correlation and is well suited to the analysis of categorical data.

Latent Gaussian Models
Our model uses latent Gaussian variables to model the distribution of categorical observations. For categorical data, each element $y_{ni}$ of the observed vector $y_n$ can take values from a finite discrete set $S_n = \{0, 1, \ldots, C_n\}$. For the n-th data vector, the generative process is: (1) Sample latent Gaussian vector $z_n \sim \mathcal{N}(0, \Sigma)$. (2) Take a linear combination of $z_n$ to obtain the predictor $x_n = Wz_n$. (3) Draw data from a categorical distribution given $x_n$. Our goal is to learn the model parameters $\theta = (\mu, \Sigma)$.

Variational Inference
Variational lower bound: Computation of the marginal likelihood is intractable since the categorical likelihood is not conjugate to the Gaussian prior. Using Jensen’s inequality, we can obtain a lower bound to the marginal likelihood:

$$L(\theta) = \sum_{i=1}^{n} \log \sum_{x_i} \frac{p(x_i|\theta)}{q(x_i)} \geq \sum_{i=1}^{n} \log \frac{\sum_{x_i} p(x_i|\theta) q(x_i)}{q(x_i)}$$

Piecewise bounds for the stick-breaking likelihood: These expectations of the likelihood w.r.t. a Gaussian, are still intractable due to the presence of $\log(1 - \exp(x))$, the logistic-log-partition (LLP) function. This is made tractable by using piecewise lower bounds for the LLP, for which the expectation of each piece w.r.t. a Gaussian is tractable.

Results
Categorical Latent Gaussian Graphical Model (CLGGM)

Highly correlated synthetic data: $|D| = 2 \times K = 4$, $N = 10,000$ We compare the true discrete distribution with the estimated distribution using various methods and models.

Real data: We compare missing value imputation error for two datasets:
- Tic-tac-toe: $K = 29$, $D = 10$, $N = 968$.
- AES-16: $K = 60$, $D = 17$, $N = 913$.

Multi-class Gaussian Process Classification

Forensic glass data: $(D = 214, K = 6)$, The top row shows contour plots of negative log-likelihood for the training data obtained under various hyperparameter settings. The bottom row shows the prediction error on the test set. The star indicates the hyperparameter setting at the minimum negative log-likelihood.

Illustrating LGM on voting dataset ($D = 16$, $K = 2$, $N = 435$). We plot the posterior mean of the latent factors with size proportional to the log-likelihood. We also plot the probability of two variables (votes) taking the same value and the probability of voting ‘yes’ given the party.