We develop data augmentation methods unique to the negative binomial (NB) distribution.

We perform joint count and mixture modeling under the NB process, using completely random measures (Poisson process, gamma process, beta process, NB process) that are simple to construct and amenable for posterior computation.

We propose to augment-and-conquer the NB process: by “augmenting” a NB process into both the gamma-Poisson and compound Poisson representations, we “conquer” the unification of count and mixture modeling, the analysis of fundamental model properties, and the derivation of efficient Gibbs sampling.

We show that the gamma-NB process can be reduced to the hierarchical Dirichlet process with normalization, highlighting its unique theoretical, structural and computational advantages.

A variety of NB processes with distinct sharing mechanisms are constructed and applied to topic modeling, with connections to existing algorithms, showing the importance of inferring both the NB dispersion and probability parameters.

**Poisson Process for Count & Mixture Modeling**

Mixture modeling infers probability random measures to assign data points into clusters (mixture components).

Since the number of points assigned to clusters are counts, we consider mixture modeling as a count-modeling problem.

The NB distribution, parameterized by a dispersion parameter and a probability parameter, is used to model overdispersed counts.

The Poisson process provides not only a way to generate independent counts from each partition of the space, but also a mechanism for mixture modeling, which allocates the observations into any measurable disjoint partition of space, conditioning on the total count and the normalized mean measure.

\[ X_j(\Omega) = \sum_{x_{ij} \in X_j} X_j(a_{ij}), \quad X_j(A_k) = \text{Pois}(G(A_k)), \]

\[ X_j(\Omega) \sim \text{Pois}(\gamma(\Omega)), \quad [X_j(A_1), \ldots, X_j(A_k)] \sim \text{Mult}(X_j(\Omega); \vec{G}(A_1), \ldots, \vec{G}(A_k)) \]

**Augment-and-Conquer the Negative Binomial Distribution**

**The joint distribution of the customer count and table count are equivalent:**

**Gamma-Negative Binomial Process**

\[ X_j \sim \text{NB}(G, \rho_j), \quad G \sim \text{Ga}(\xi, \eta), \]

\[ X_j(\Omega) \sim \text{Pois}(G(\Omega)), \quad X_j(A_k) \sim \text{Pois}(G(A_k)), \]

\[ X_j(\Omega) \sim \text{Pois}(G(\Omega)), \quad X_j(A_k) \sim \text{Pois}(G(A_k)) \]

\[ \text{Posterior Analysis} \]

\[ \Lambda_j(\Omega), X_j(\Omega) \sim \text{Beta}(a_0 + X_j(\Omega), b_0 + G(\Omega)) \]

**Predictive Distribution**

\[ X_j(\Omega) | G \sim \text{Beta}(a_0 + X_j(\Omega), b_0 + G(\Omega)) \]

**Related to the Hierarchical Dirichlet Process**

\[ X_j(\Omega) \sim \Lambda_j(\Omega), \quad \Lambda_j \sim \text{DP}(a, G), \quad a \sim \text{Beta}(\alpha_0, \beta_0), \quad G \sim \text{Dir}(\alpha_0) \]

The NB process family and related algorithms are compared in the table below.

**Table 1:** Comparison of per-word perplexities on the held-out words between various algorithms. (a) With 60% of the words in each document used for training, the performance varies as a function of \( \alpha \), which is set to 0.01 in both LDA and NB-LDA, which are parametric models, whereas the NB-NB, NB-FTM, Beta-NB, CRF-NB, Gamma-NB, and Marked-Beta-NB all infer the number of active topics, which are 127, 201, 105, 163, 177 and 130, respectively, according to the last Gibbs sampling iterations. (b) Per-word perplexities of various models as a function of the percentage of words in each document used for training. The results of the LDA and NB-LDA are shown with the best settings of \( \alpha \) under each training/testing partition.

**Comparison of Per-Word Perplexities**

**Inferred NB Process Model Parameters**

**Figure 1:** Comparison of per-word perplexities on the held-out words between various algorithms. (a) With 60% of the words in each document used for training, the performance varies as a function of \( \alpha \), which is set to 0.01 in both LDA and NB-LDA, which are parametric models, whereas the NB-NB, NB-FTM, Beta-NB, CRF-NB, Gamma-NB, and Marked-Beta-NB all infer the number of active topics, which are 127, 201, 105, 163, 177 and 130, respectively, according to the last Gibbs sampling iterations. (b) Per-word perplexities of various models as a function of the percentage of words in each document used for training. The results of the LDA and NB-LDA are shown with the best settings of \( \alpha \) under each training/testing partition.