Augment-and-Conquer Negative Binomial Processes

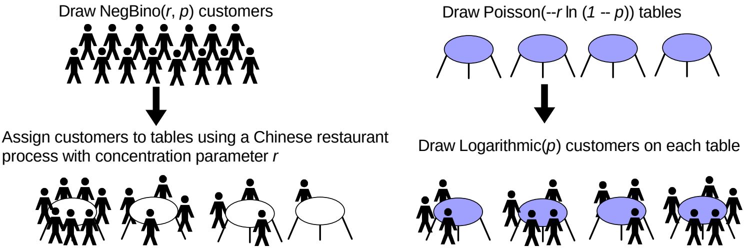
Mingyuan Zhou and Lawrence Carin

Department of Electrical & Computer Engineering
Duke University, Durham, NC 27708, USA

NIPS, Lake Tahoe, December 05, 2012

- We view mixture modeling as a count-modeling problem
  - Number of words assigned to topic k in document j (we are modeling a $K \times J$ count matrix in a mixed-membership model)
- The NB distribution $x \sim \text{NB}(r, p)$ can be augmented as
  - a gamma-Poisson mixture distribution
  - a compound Poisson distribution
- We discover a Poisson-logarithmic bivariate count distribution

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



- The NB process $X \sim \mathrm{NBP}(G_0, p)$ augmented as $X \sim \sum_{t=1}^{L} \mathrm{Log}(p),\ L \sim \mathrm{PP}(-G_0 \ln(1-p))$ is equivalent in distribution to $L \sim \mathrm{CRTP}(X, G_0),\ X \sim \mathrm{NBP}(G_0, p)$.
- Gamma-NB Process: $X_j \sim \mathrm{NBP}(G, p_j),\ G \sim \mathrm{GaP}(c, G_0)$
- Beta-NB Process: $X_j \sim \mathrm{NBP}(r_j, B),\ B \sim \mathrm{BP}(c, B_0)$
- Marked-BNBP: $X_j \sim \mathrm{NBP}(R, B),\ (R, B) \sim \mathrm{MBP}(c, G_0, B_0)$

Table: The negative binomial process family: $X_j(\omega_k) \sim \mathrm{NB}(r_{(j,k)}, p_{(j,k)})$.

Algorithms	r_k	r_j	p_k	p_j	π_k	Related Algorithms
NB-LDA		✓		✓		NMF, LDA, Dir-PFA
NB-HDP	✓			0.5		HDP, DILN-HDP
NB-FTM	✓			0.5	✓	FTM, S$\gamma\Gamma$-PFA
Beta-NB		✓	✓			Beta-Geometric, BNBP
Gamma-NB	✓			✓		CRF-HDP
Marked-Beta-NB	✓		✓			BNBP

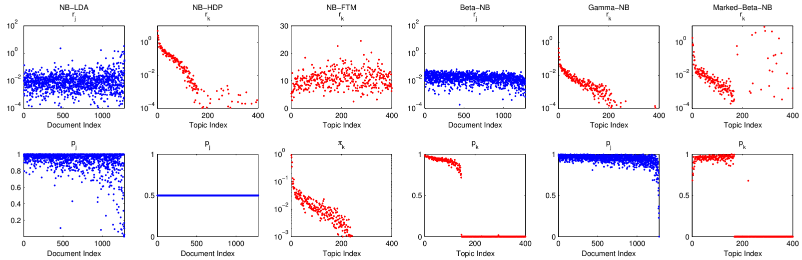


Figure: Distinct sharing mechanisms and model properties are evident between various NB processes, by comparing their inferred parameters.

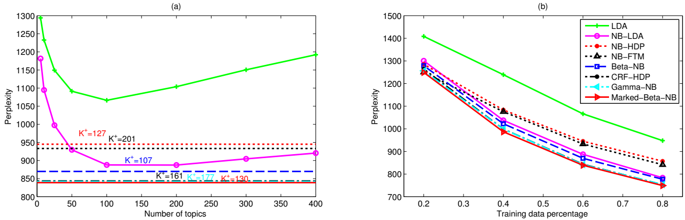


Figure: Comparison of per-word perplexities between various algorithms.