Indian Buffet Process with Power-law Behaviour

Bobby He, Hector McKimm, Natalia Garcia Martin

October 18, 2019

Bayesian Nonparametric (BNP) Models

- Model selection crucial in statistical problems
- Featual models:
  - How to choose number of features?
- BNP models: number of features allowed to grow with observations
- Number of features unbounded
- Observe finite number of features

Indian Buffet Process (IBP)

- Stochastic process to generate an infinitely exchangeable distribution over infinite binary matrices.
- Observe objects $Z_i, i = 1, \ldots, n$
  - Represent objects as binary vectors
- Binary matrix Z with rows Z_i

- The first customer tries Poisson(α) dishes.
- The n^{th} customer tries:
  - Each of the current dishes with probability $m_k/c + n - 1$, where m_k denotes the number of times dish k has been chosen, and
  - Poisson$(\alpha c/c + n - 1)$ new dishes.

Ghahramani and Griffiths ((2006))

Two-parameter generalization of the Indian Buffet Process

$$B \sim \mathrm{BP}(c, B_0) \tag{1}$$

$$Z_i \mid B \sim \mathrm{BernoulliP}(B) \tag{2}$$

Can be shown that:

$$Z_1 \sim \mathrm{BernoulliP}(B_0). \tag{3}$$

$$Z_{n+1} \mid Z_1, \ldots, Z_n \sim \mathrm{BernoulliP}\left(\frac{c}{c+n} B_0 + \sum_j \frac{m_j}{c+n} \delta_{\theta_j^*}\right). \tag{4}$$

B_0 a continuous measure.
θ_j^* the unique atoms in $Z_1, \ldots, Z_n$ with θ_j^* appearing m_j times

Thibaux and Jordan ((2007))

Stable-beta Process

- The SBP is a CRM with no fixed atoms and Lévy measure

$$\Lambda_0(dp, d\theta) = \alpha \frac{\Gamma(1+c)}{\Gamma(1-\sigma)\Gamma(c+\sigma)} p^{-\sigma-1}(1-p)^{c+\sigma-1} dp H(d\theta).$$

- Parameters $\alpha > 0$, $\sigma \in [0,1)$ and $c > -\sigma$.
- 3-parameter generalisation of IBP defined by:

$$\begin{aligned} \mu &\sim \mathrm{CRM}(\Lambda_0, \{\}), \\ Z_i \mid \mu &\sim \mathrm{BernoulliP}(\mu). \end{aligned}$$

Stable-beta IBP

- The first customer tries Poisson(α) dishes.
- The n^{th} customer tries:
  - dish k with probability $\frac{m_k - \sigma}{c+n-1}$, where m_k is popularity of dish k.
  - as well as Poisson$\left(\alpha \frac{\Gamma(1+c)\Gamma(n-1+c+\sigma)}{\Gamma(n+c)\Gamma(c+\sigma)}\right)$ new dishes.
- Generalisation of 2-parameter IBP ($\sigma = 0$)
- Parameter Interpretation:
  - $\alpha =$ mean number of dishes for each customer
  - c and σ control propensity to choose new dishes
  - σ is the limiting proportion of dishes chosen by 1 customer

Power-law behaviour

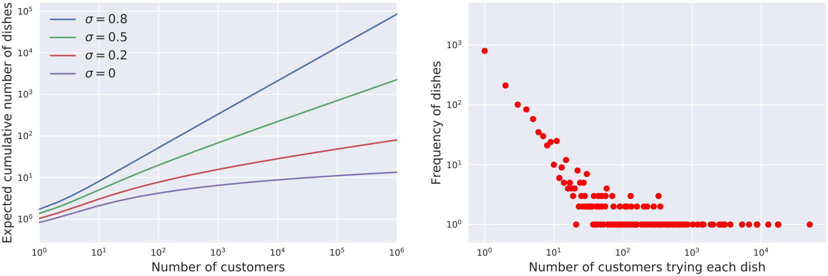

Figure: Power-law behaviour of the SB-IBP (log-log scale).

20newsgroups dataset

- Collection of approximately 20,000 news documents partitioned across 20 newsgroups
- Documents = customers, words = dishes
- Recreate results from Teh and Görür ((2009))
- Fit ML parameters to each newsgroup for both IBP and Stable-beta IBP
- Compare goodness-of-fit

20newsgroups dataset

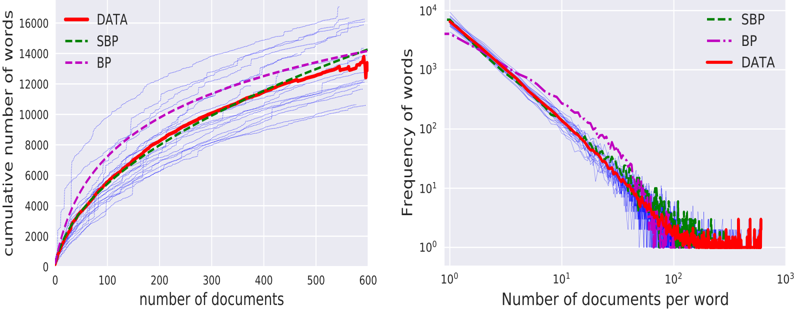

Figure: Power-law properties of the 20newsgroups dataset.

NeurIPS data

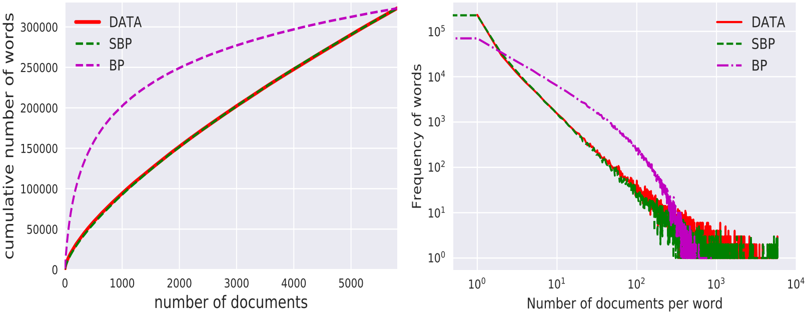

Figure: Power-law properties of the Neurips dataset.

References

Z. Ghahramani and T. L. Griffiths. Infinite latent feature models and the indian buffet process. pages 475–482, 2006.

Y. W. Teh and D. Görür. Indian buffet processes with power-law behavior.
pages 1838–1846, 2009. URL `http://papers.nips.cc/paper/3638-indian-buffet-processes-with-power-law-behavior.pdf`.

R. Thibaux and M. I. Jordan. Hierarchical beta processes and the indian buffet process. In *Artificial Intelligence and Statistics*, pages 564–571, 2007.