

# Nonparametric Bayesian Statistics

Debdeep Pati  
Florida State University

October 20, 2014

- **Pitman-Yor Processes** are also known as **Two-parameter Poisson-Dirichlet Processes**.
- Chinese restaurant representation:

$$P(n^{\text{th}} \text{ customer sit at table } k, 1 \leq k \leq K) = \frac{n_k - d}{n - 1 + \alpha}$$

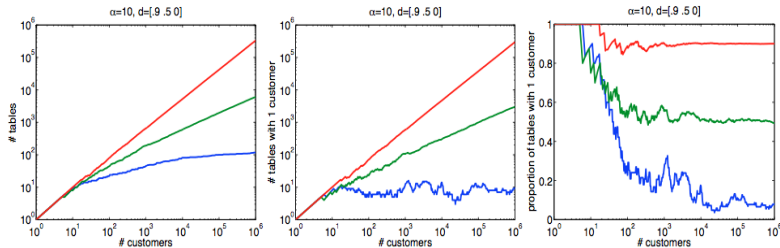
$$P(n^{\text{th}} \text{ customer sit at new table}) = \frac{\alpha + dK}{n - 1 + \alpha}$$

where  $0 \leq d < 1$  and  $\alpha > -d$ .

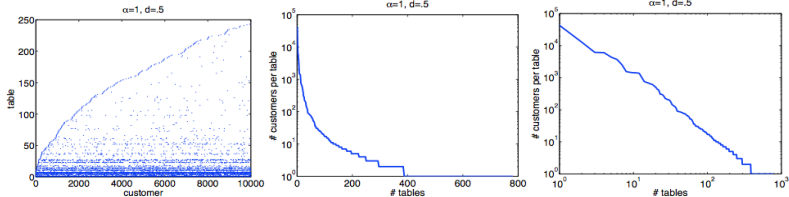
- When  $d = 0$  the Pitman-Yor process reduces to the DP.

# Pitman-Yor processes

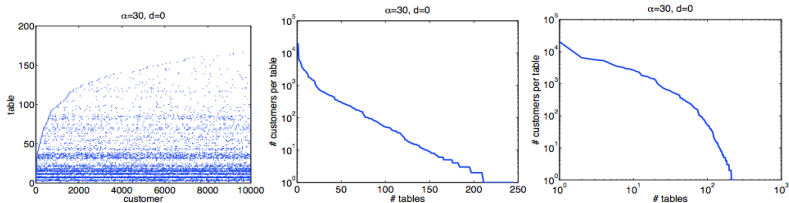
- Two salient features of the Pitman-Yor process:
  - With more occupied tables, the chance of even more tables becomes higher.
  - Tables with smaller occupancy numbers tend to have lower chance of getting new customers.
- The above means that Pitman-Yor processes produce Zipf's Law type behaviour.



## Draw from a Pitman-Yor process



## Draw from a Dirichlet process



- We can relax the priors on  $\beta_k$  in the stick-breaking construction:

$$G = \sum_{k=1}^{\infty} \pi_k \delta_{\theta_k^*} \qquad \pi_k = \beta_k \prod_{l=1}^{k-1} (1 - \beta_l)$$
$$\theta_k^* \sim H \qquad \beta_k \sim \text{Beta}(a_k, b_k)$$

- We get the DP if  $a_k = 1, b_k = \alpha$ .
- We get the Pitman-Yor process if  $a_k = 1 - d, b_k = \alpha + kd$ .
- To ensure that  $\sum_{k=1}^{\infty} \pi_k = 1$ , we need  $\beta_k$  to not go to 0 too quickly:

$$\sum_{k=1}^{\infty} \pi_k = 1 \quad \text{almost surely iff} \quad \sum_{k=1}^{\infty} \log(1 + a_k/b_k) = \infty$$

- ▶ One possibility for a two group DDP is to let

$$f(y | x) = \sum_{h=1}^{\infty} \pi_h \mathbf{N}(y; \mu_{xh}, \tau^{-1})$$
$$\mu_{1h} \sim \pi \delta_{\mu_{0h}} + (1 - \pi) \delta_{\beta_h},$$
$$\mu_{0h} \sim N(\mu, \sigma_{\mu}^2), \beta_h \sim N(\mu_{\beta}, \sigma_{\beta}^2),$$

- ▶ One can test for equality of distributions.