# Approximated Ground Truth

**Donglai Wei**

2010.5.28

## 1) Approximated Ground Truth

So, what is the Ground Truth? Maybe 11 dishes config is better than 10 real bars?
Approximation of the Ground Truth:

1. Initialization:

    (a) Given the real bars, we can easily divide each restaurant(noticably,there exists some noisy customers)

    (b) For the overlapped part, we first distribute the customers equally
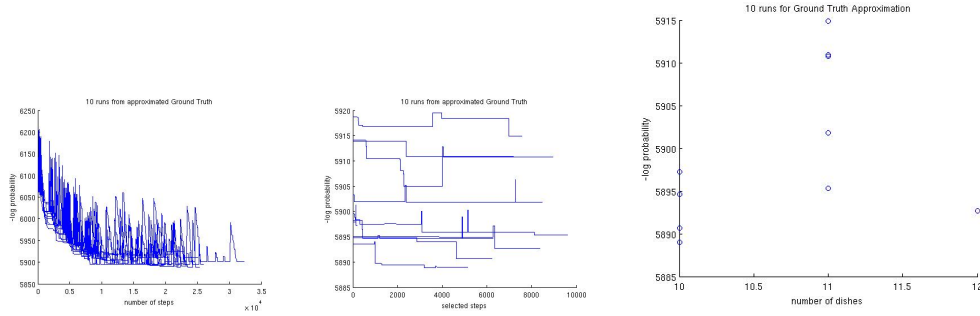
2. Run the search algorithm combo described above 10 times



Figure 1: Approximated Ground Truth:left,ten plots against steps;middle,select part of left;right,"10 dishes" is favorable

## 2) Formula

Maximize the log likelihood:

$P = -log p(x, z|\lambda)$

$=$

$(\text{t-term}) log \frac{\Gamma(m_{..}+\gamma)}{\Gamma(\gamma)} + \sum_{j=1}^{J} \{ log \frac{\Gamma(n_{j..}+\alpha)}{\Gamma(\alpha)} - \sum_{t=1}^{m_j} [log(\Gamma(n_{jt.}) + log\alpha] \}$

$+(\text{k-term}) \sum_{k=1}^{K} [log(\frac{\Gamma(n_{..k}+W\phi_0)}{\Gamma(W\phi_0)}) + log(\Pi_{w=1}^{W} \frac{\Gamma(\phi_0)}{\Gamma(\phi_0+n_{..k}^{w})}) - log(\Gamma(m_{.k}) - log\gamma]$

1

W:number of unique words

$n^w_{..k}$number of occurence of word w in dish k