Purpose

- These slides derive a quantity proportional to the posterior probability of the topic assignments.

- Tracking this quantity will help in assessing the convergence of a collapsed Gibbs sampler for LDA to the desired stationary distribution.

Graphical Model with Generative Story

For each topic $i$:
- Sample a distribution over words.
- For each document $j$:
  - Sample a distribution over topics.
  - Choose a document length $M_j$ (tokens).
  - For each position $j$ in document $i$:
    - Sample a topic.
    - Sample a word token from that topic.

Joint Likelihood (2)

\[
P(w \mid z, \alpha, \beta) = \prod_{i=1}^{n} P(z_i \mid \alpha) \cdot \prod_{j=1}^{M} P(w_j \mid z_j, \beta)
\]

- Focusing on the first term:
  \[
P(z_i \mid \alpha) = \prod_{i=1}^{n} \prod_{c=1}^{C} \frac{\Gamma(\alpha_c)}{\Gamma(\alpha)} \left( \sum_{c=1}^{C} \frac{\alpha_c}{\sum_{c=1}^{C} \alpha_c} \right)^{\alpha_c}
  \]

- Focusing on the second term:
  \[
P(w_j \mid z_j, \beta) = \prod_{j=1}^{M} \prod_{w=1}^{V} \frac{\Gamma(\beta_{wz_j})}{\Gamma(\beta)} \left( \sum_{w=1}^{V} \frac{\beta_{wz_j}}{\sum_{w=1}^{V} \beta_{wz_j}} \right)^{\beta_{wz_j}}
  \]

Joint Likelihood (3)

- \(P(z_i \mid \alpha)\) is the Beta function, and \(\Gamma()\) is the Gamma function.
- \(\alpha_c\) is the number of times topic $i$ is assigned to tokens in document $d$.
- \(\alpha^{(i)}\) is the vector whose elements are $\alpha_i^{(t)}, 1 \leq t \leq T$.
- $g$ is the length $T$ symmetric vector $[a, a, ..., a]$.
- \(\beta_{wz_j}\) is the vector whose elements are $\beta_{wz_j}, 1 \leq w \leq V$.
- $\beta$ is the length $V$ symmetric vector $[\beta, \beta, ..., \beta]$.
Joint Likelihood (4)

- Putting the parts together:
  \[ p(y \mid \theta) = \prod_{i=1}^{T} \left[ \prod_{j=1}^{K} \left( \frac{1}{\theta_j + x_j} \right)^{y_{ij}} \right] \]
  
- After removing constants, the quantity to track is:
  \[ r(y \mid \theta) = \prod_{i=1}^{T} \left[ \prod_{j=1}^{K} \left( \frac{1}{\theta_j + x_j} \right)^{y_{ij}} \right] \]

- In practice, the logarithm of this quantity will be easier to manipulate.

- Plot the log-quantity at every iteration of the sampler to assess convergence.

**Variables:**
- \( \theta \) is the vector whose elements are \( \theta_i, 1 \leq i \leq T \)
- \( x \) is the length-\( T \) symmetric vector \( [x, \ldots, x] \)
- \( y_{ij} \) is the vector whose elements are \( y_{ij}, 1 \leq i \leq T \)
- \( z \) is the length-\( T \) symmetric vector \( [z, \ldots, z] \)