(i) \( p(\lambda_1, \lambda_2 | y) \propto p(y | \lambda_1, \lambda_2) p(\lambda_1) p(\lambda_2) \)

Since \( p(\lambda_1, \lambda_2 | y) = p(y | \lambda_1, \lambda_2) \) (can be factored), we know \( \lambda_1, \lambda_2 \) are independent.

\[ p(\lambda_1 | y) = \text{Gamma}(\lambda_1 | a_i + \sum_{k \neq i} y_k, b_i + k) \]
\[ p(\lambda_2 | y) = \text{Gamma}(\lambda_2 | a_l + \sum_{k \neq i} y_k, b_l + (T - k)) \]

Note: Several people tried to do i. by finding \( p(\lambda, y) \propto p(y | \lambda_1) p(\lambda_1) \) but \( y \)'s dist depends on both \( \lambda_1, \lambda_2 \) so you can't do this calculation. \( p(\lambda_1 | y) \) can be found from the joint dist given above.

(ii) \( p(\lambda_1, \lambda_2, k | y) \propto p(y | \lambda_1, \lambda_2, k) p(\lambda_1) p(\lambda_2) p(k) \)

\[ \propto \left( \prod_{i=1}^{k} \lambda_{i_k}^{-y_i} e^{-\lambda_i} \right) \left( \prod_{i=1}^{k} \lambda_{i_k}^{-y_i} e^{-\lambda_2} \right) \lambda_1^{-a_i - \lambda_1} \lambda_2^{-a_l - \lambda_2} e^{-k} \]

(Same as before but \( p(k) \) is uniform)

(iii) Gibbs sampling: \( p(\lambda_1, \lambda_2 | y, k) \) is the posterior dist from \( \lambda_1, \lambda_2 \) hence known.

\( p(k | \lambda_1, \lambda_2, y) \) is not a known dist but it is a discrete dist taking values in \( \{1, 2, \ldots, T - 1\} \). Just compute above at each value and renormalize to create distribution.

Note: A grid here would be 3-dimensional. Probably not practical if \( T \) is large.

(c) i. \( p(y_t | y_{-t}) = \int p(y_t | \lambda_1, y_{-t}) p(\lambda_1 | y_{-t}) d\lambda_1 \)

\[ = \int p(y_t | \lambda_1) p(\lambda_1 | y_{-t}) d\lambda_1 \]

(\( y_t \) is indep of \( Y_{-t} \) given \( \lambda \))

ii. Compute \( p(y_t | y_{-t}) \). If this is small then \( y_t \) is unusual given the other data. Of course it can be hard to judge "small" if order of magnitude of \( y_t \) is large because then every value is unlikely. In such a case we might sample \( y_t^{(n)} \) from \( p(y_t | y_{-t}) \) and see where observed \( y_t \) falls in the distribution.

2. In this case \( Y_{ij} \) are not iid given \( \theta_i \) since they also depend on \( M_{ij} \). They are not exchangeable. If all of the \( M_{ij} \)'s for a given \( i \) are equal, then the \( Y_{ij} \) will be exchangeable.
\( p(\theta, \alpha, \beta | Y) \propto \prod_{i=1}^{T} \prod_{j=1}^{n_i} \theta_i (\theta_i^j)^{Y_{ij}} e^{-\theta_i^j (\beta + \bar{X} m_j Y_{ij})} \) 

\( p(\theta | \alpha, \beta, Y) \propto \prod_{i=1}^{T} \theta_i^{\alpha + n_i - 1} e^{-\theta_i (\beta + \bar{X} m_j Y_{ij})} \) (only include terms depending on \( \theta \))

\( \Rightarrow \theta_i, i=1, \ldots T \) are independent Geo(\( \alpha + n_i, \beta + \bar{X} m_j Y_{ij} \)) r.v.

3. The conditional posterior distribution takes the prior info (\( \alpha, \beta \)) and updates it to include the data yielding \( \theta_i (\alpha + n_i, \beta + \bar{X} m_j Y_{ij}) \). Another way to see this is to look at prior and posterior means:

\[
\text{prior mean of } \theta_i = \frac{\alpha}{\beta}, \quad \text{posterior mean } = \frac{\alpha + n_i}{\beta + \bar{X} m_j Y_{ij}} = \frac{\alpha}{\beta} + \frac{n_i}{\beta + \bar{X} m_j Y_{ij}} \sum_{j} m_j Y_{ij}
\]

which combines prior mean \( \frac{\alpha}{\beta} \) and MLE for \( \theta_i \frac{n_i}{\sum_{j} m_j Y_{ij}} \)

4. This is just like the HW and example in class.

\textbf{Approach 1:} Find \( p(\alpha, \beta | Y) = \frac{p(\alpha, \beta, \theta | Y)}{p(\theta | \alpha, \beta, Y)} \) from 6

- Use 2-D grid approx to \( p(\alpha, \beta | Y) \) to generate \( (\alpha, \beta) \) sample
- Sample \( \theta \) from \( p(\theta | \alpha, \beta, Y) \)

\textbf{Approach 2: MCMC} — We know \( p(\theta | \alpha, \beta, Y) \)

- Need to say something about \( p(\alpha, \beta | \theta, Y) \) (HARD!)

5. Given posterior sample, for \( \theta_i \), say \( \theta_i^{(1)}, \theta_i^{(2)}, \ldots, \theta_i^{(M)} \) (from part 6), we just sample \( y_{\text{new}} \sim \text{Exp}(\theta_i^{(K)} \cdot 2) \) \( k=1, \ldots, M \).

6. Here we need \( y_{\text{new}} \). To simulate the predictive distribution.

\textbf{For } k=1, \ldots, M

- Generate \( \theta_{\text{new}} \sim p(\alpha, \beta | Y) \) (a new job type parameter)
- Generate \( y_{\text{new}} \sim \text{Exp}(\theta_{\text{new}} \cdot 2) \)