Question 6.

Suppose we observe pairs (Y_i, X_i), where $Y_i \sim \text{Poisson}(m\beta\tau_i)$, $i = 1, \cdots, n$ are mutually independent random variables, and $\mathbf{X} = (X_1, \cdots, X_n) \sim \text{Multinomial}(m; \boldsymbol{\tau})$, $\boldsymbol{\tau} = (\tau_1, \tau_2, \cdots, \tau_n)$ with $\sum_{i=1}^n \tau_i = 1$. Suppose $\mathbf{X}$ and $\mathbf{Y} = (Y_1, \cdots, Y_n)$ are independent. Think of Y_i as the incidence of a disease in region i, where the underlying rate is $\beta\tau_i$, a multiplicative factor of an overall effect β, where τ_i is a regional factor, a measure of population density in region i for example. The factor τ_i is not observed, but information on τ_i is obtained using X_i. Treat $m = \sum_{i=1}^n x_i$ as known.

1. Find the joint density of $\mathbf{Y} = (Y_1, \cdots, Y_n)$ and $\mathbf{X} = (X_1, \cdots, X_n)$. (**2 marks**)
   Solution:

2. Find the MLEs of β and τ_i. (**5 marks**)

Solution:

3. Suppose that x_1 is missing. Use the fact that $X_1 \sim \text{Binomial}(m, \tau_1)$ to calculate the expected complete-data log likelihood. Find the EM sequence, i.e. $\hat{\beta}^{(r)}$ and $\hat{\tau}_i^{(r)}$ for r^{th} iteration of the EM algorithm. **(8 marks)**

   <u>**Solution:**</u>