COMPREHENSIVE WRITTEN EXAMINATION, PAPER III

FRIDAY AUGUST 17, 2007, 9:00 A.M.–1:00 P.M.

STOR 665 QUESTIONS

Summer Interns at Bank of America (BoA)

Suppose you are one of the group of summer interns at the Statistical Decision Support group at BoA's headquarter in Charlotte, NC. On your first day at work, your supervisor Mr. Bing assigns the following project to you.

BoA has one of the largest customer service call centers in the world. Call centers are the main service channel that BoA customers can speak with banking representatives. To serve customers well has always been BoA's #1 priority. To achieve a higher quality of service, the management needs to assign enough reps, which follows from an accurate forecast of future call arrival rates. Sometimes, a few hours into a business day, the management wants to update the previous rate forecast using the call volumes observed in the first few hours of the day. You now have the opportunity to develop statistical models to provide the management with such forecasts. You will compete with each other and the winner will get a pair of tickets to Charlotte Bobcats' opening game. Work Hard!

To help you get started, Mr. Bing is kind enough to provide the following background information. For a particular day, the business period (for example, 7am to midnight) is divided into 15-minute intervals, and the number of calls arrived during each interval will be recorded. People have been modelling such numbers as observations from Poisson random variables with their rates depending on the day and the interval.

To be consistent, here are some notations. Let $\mathbf{Y} = (y_{ij})$ be an $n \times m$ matrix that records the arrival counts from n consecutive days with each day being aggregated into m time intervals. Furthermore, we assume that y_{ij} is a Poisson random variable with rate λ_{ij}. Let $\boldsymbol{\Lambda} = (\lambda_{ij})$ denote the $n \times m$ Poisson rate matrix. The ith row of $\boldsymbol{\Lambda}$, denoted as $\boldsymbol{\lambda}_{(i)}^T = (\lambda_{i1}, \ldots, \lambda_{im})$, is the *rate profile* of the ith day. Correspondingly, the ith row of $\mathbf{Y}$, denoted as $\mathbf{y}_{(i)}^T = (y_{i1}, \ldots, y_{im})$, is the *count profile* of the ith day.

1. **[50 points]** Consider the following factor model for the count matrix $\mathbf{Y}$:

$$\begin{cases} \mathbf{y}_{(i)} \sim \text{Poisson}(\boldsymbol{\lambda}_{(i)}), \quad i = 1, \ldots, n, \\ g(\boldsymbol{\lambda}_{(i)}) = \beta_{i1}\mathbf{f}_1 + \cdots + \beta_{iK}\mathbf{f}_K \equiv \mathbf{F}\boldsymbol{\beta}_{(i)}, \end{cases} \tag{1}$$

where $\mathbf{F}_{m \times K} = (\mathbf{f}_1, \ldots, \mathbf{f}_K)$ contains K underlying factors in $\mathbb{R}^m$, $\boldsymbol{\beta}_{(i)} = (\beta_{i1}, \ldots, \beta_{iK})^T$ is the factor loading vector for the ith rate profile.

(a) **[10 points]** Suppose you are interested in day $n + 1$ and the management will perform an updating at the end of time period m_0. Identify from the project description the two forecasting problems that you need to solve. Describe them using the above mathematical notations.

(b) **[10 points]** What is the benefit for forecasting $\boldsymbol{\lambda}_{(n+1)}$ when using the above factor model? Justify your answer. Note that in practice one would usually use a small K.

(c) **[10 points]** What is the statistical name of the tranformation g? What is the most common choice for g?

(d) **[20 points]** The model (1) can be estimated using maximum likelihood method. Assume $K = 1$. Direct estimation of the factor model may be complicated because both $\boldsymbol{\beta}$ and $\mathbf{f}$ are unknown.

Describe an iterative algorithm for model estimation. Can you show whether the algorithm converges?

2. **[50 points]** Consider the within-day updating problem. In addition to the historical count profiles $\{\mathbf{y}_{(1)}, \ldots, \mathbf{y}_{(n)}\}$, suppose you also observe the counts during the first m_0 time periods of the $(n+1)$th day. Denote the new observations collectively as $\mathbf{y}^e_{(n+1)} = (y_{n+1,1}, \ldots, y_{n+1,m_0})^T$, and the corresponding Poisson rate vector as $\boldsymbol{\lambda}^e_{(n+1)}$. Let $\mathbf{y}^l_{(n+1)} = (y_{n+1,m_0+1}, \ldots, y_{n+1,m})^T$ be the latter count profile with the rate vector being $\boldsymbol{\lambda}^l_{(n+1)}$.

   Suppose you have estimated the K-factor model using the n historical count profiles, and denote the K factors as $\mathbf{f}_1, \ldots, \mathbf{f}_K$ and their loadings as $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K$. Assume somehow you have derived a forecast for the loading vector of the $(n+1)$th day, $\tilde{\boldsymbol{\beta}}_{(n+1)} = (\tilde{\beta}_{(n+1),1}, \ldots, \tilde{\beta}_{(n+1),K})$.

   To obtain an updated forecast of $\boldsymbol{\beta}_{(n+1)}$ after observing $\mathbf{y}^e_{(n+1)}$, you propose to minimize the following criterion with respect to $\boldsymbol{\beta}_{(n+1)}$,

$$\sum_{j=1}^{m_0} \left[\lambda_{n+1,j} - y_{n+1,j} \log\left(\lambda_{n+1,j}\right)\right] + \omega \sum_{k=1}^{K} \left(\beta_{n+1,k} - \tilde{\beta}_{n+1,k}\right)^2, \qquad (2)$$
$$\text{subject to} \quad g(\boldsymbol{\lambda}^e_{(n+1)}) = \mathbf{F}^e \boldsymbol{\beta}_{(n+1)},$$

   where $\mathbf{F}^e$ is an $m_0 \times K$ matrix whose (j,k)th entry is f_{jk}, $1 \le j \le m_0$, $1 \le k \le K$.

   (a) **[10 points]** The above criterion (2) makes a lot of sense to Mr. Bing, a statistician. However, you will need to explain it to the upper management using plain English. Here are some hints.

   - The criterion consists of two parts. What are they? How do you explain them?
   - Why do you sum up the two parts?

   (b) **[10 points]** What is the function of the extra parameter ω? What will happen if $\omega = 0$? How about $\omega \to \infty$? Can you think of some data-driven approaches to select some "optimal" value for it?

   (c) **[30 points]** Direct minimization of (2) seems daunting. One possible approach is to derive some iterative reweighted least squares algorithm. That means trying to find some quadratic approximation of (2) around some initial estimate $\boldsymbol{\beta}^0$, solve the quadratic criterion, and then iterate until the solution converges.

   Derive such an algorithm for the commonly used function g you identified earlier in Question 1(c).