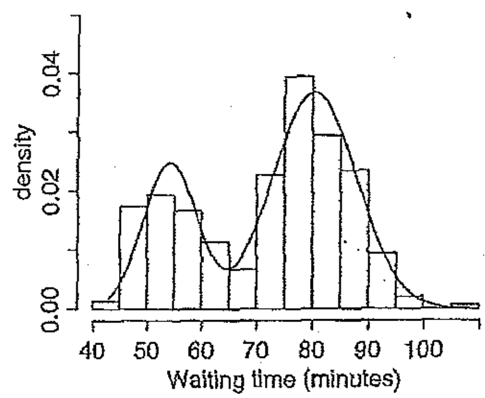
COMPREHENSIVE WRITTEN EXAMINATION, PAPER III

FRIDAY AUGUST 15, 2014, 9:00 A.M.–1:00 P.M.

STOR 665 QUESTIONS

1. **[40 points]**

   The following histogram is about the waiting time (in minutes) of $N = 299$ consecutive eruptions of the Old Faithful geyser in Yellowstone National Park. The bimodal pattern suggests that the waiting times come from two populations, each of which can be modeled using a normal distribution.



   Let Y_i denote the ith waiting time, $i = 1, \dots, 299$. A suitable model for the data is the following two-component normal mixture model:

$$Y_i = (1 - \delta_i) \cdot Y_{i1} + \delta_i \cdot Y_{i2},$$

   where δ_i is a random sample from the Bernoulli distribution with success probability π, Y_{i1} is a random sample from $N(\mu_1, \sigma_1^2)$, and Y_{i2} is a random sample from $N(\mu_2, \sigma_2^2)$. In addition, δ_i, Y_{i1}, Y_{i2} are mutually independent. Denote the parameters as $\theta = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \pi)^T$.

   Your task is to come up with a suitable estimator for the parameters θ. You need to clearly state the motivation for your estimation method, and derive an explicit estimation algorithm.

2. **[60 points]** The following exercise is about understanding customer patience while waiting to be served by a telephone call center agent. Imagine the following scenario: you are calling your bank for some banking needs; you are listening to some bank commercials, while waiting for an agent to become available; while waiting, two things can happen - either your patience runs out and you hang up the phone before being served; or you are lucky enough to get served before losing patience.

The amount of time you actually waited, W, can serve as a proxy for your patience. If you hang up before service, your patience equals W; otherwise, you are served and we know that your patience is longer than W, but not sure how much longer. In the latter case, we say that the observation W is censored, with the censoring indicator $\delta = 1$. (Note that $\delta = 0$ if the observation is not censored, as in the former case.)

Suppose we have collected waiting times of n customers, denoted as W_i with the corresponding censoring indicator as δ_i, $i = 1, \ldots, n$.

(a) **[20 points]** Imagine that the patience times of the customers, X_i, are i.i.d. and follow an exponential distribution with the following density function:

$$f(x) = \theta \, e^{-\theta \, x}.$$

  i. Write down the likelihood function of θ based on the data $\{W_i, \delta_i\}$, $i = 1, \ldots, n$.

  ii. Derive the maximum likelihood estimator of θ.

  iii. Is there an intuitive interpretation of the MLE?

  iv. Discuss the implication of the exponential patience assumption.

(b) **[40 points]** We now extend the above problem in the following manner. Suppose all the waiting times are observed within the interval $[0, T]$, which are further divided into p sub-intervals:

$$0 = \tau_0 < \tau_1 < \cdots < \tau_p = T.$$

Furthermore, we assume that the parameter θ is now a piecewise constant function on $[0, T]$ in that

$$\theta(t) = \theta_j, \qquad \text{for} \quad t \;\in\; (\tau_{j-1}, \tau_j], \qquad j = 1, \ldots, p.$$

Consider the following transformed data, $i = 1, \ldots, n; j = 1, \ldots, p$:

$$W_{ij} = \begin{cases} 0, & W_i \le \tau_{j-1} \\ W_i - \tau_{j-1}, & \tau_{j-1} < W_i \le \tau_j \\ \tau_j - \tau_{j-1}, & W_i > \tau_j \end{cases} \quad \text{and} \quad \delta_{ij} = \delta_i \mathbb{I}(\tau_{j-1} < W_i \le \tau_j). \quad (1)$$

  i. Derive the likelihood function of $\theta(t)$ based on the transformed data (1), along with the corresponding MLE for $\theta(t)$.

  ii. Show that the above likelihood function is equivalent to the one using the original data $\{W_i, \delta_i\}$, $i = 1, \ldots, n$.