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 $i$th waiting time, $i = 1, \ldots, 299$. A suitable model for the data is the following two-component normal mixture model:

$$Y_i = (1 - \delta_i) \cdot Y_{1i} + \delta_i \cdot Y_{2i},$$

where $\delta_i$ is a random sample from the Bernoulli distribution with success probability $\pi$, $Y_{1i}$ is a random sample from $N(\mu_1, \sigma_1^2)$, and $Y_{2i}$ is a random sample from $N(\mu_2, \sigma_2^2)$. In addition, $\delta_i$, $Y_{1i}$, $Y_{2i}$ 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 $\theta$. 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 \( \delta_i \), \( i = 1, \ldots, n \).

(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 \( \theta \) based on the data \( \{W_i, \delta_i\}, i = 1, \ldots, n \).
ii. Derive the maximum likelihood estimator of \( \theta \).
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 \( \theta \) is now a piecewise constant function on \([0, T]\) in that

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

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

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

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 \( \{W_i, \delta_i\}, i = 1, \ldots, n \).