A new technology is developed to accurately measure the tumor cell secretion level in cancer patients. However, even so, there is a lower limit detection bound, denoted by \( \tau \), a positive constant; that is, if the level is less than \( \tau \), the technology will not be able to measure it. Suppose that we take measurements from \( n \) i.i.d patients and find that \( m \) patients of them have detectable levels with values \( X_1, \ldots, X_m \). For the rest of \( (n - m) \) patients, their levels \( X_{m+1}, \ldots, X_n \) are below \( \tau \) so undetectable. Moreover, we assume that each \( X_i, \ i = 1, \ldots, n \), follows an exponential distribution with rate \( 1/\lambda \), i.e., the density is \( \lambda e^{-x/\lambda}, x > 0 \). Our goal is to estimate \( \lambda \).

**Likelihood function**

1(a) (5 points) The observed data can be expressed as

\[ X_1, \ldots, X_m, I(X_{m+1} < \tau), \ldots, I(X_n < \tau). \]

Write down the observed likelihood function of \( \lambda \) based on these observations.

**Method of moments**

2(a) (5 points) Note that \( m \), the number of the patients whose levels are detectable, is random. What is the distribution of \( m \)?

2(b) (5 points) Show \( E[m] = ne^{-\tau/\lambda} \). Thus, a simple estimator for \( \lambda \), denoted by \( \hat{\lambda}_1 \), can be obtained by solving the following equation

\[ m = ne^{-\tau/\lambda}. \]

Show \( \hat{\lambda}_1 \) is consistent and derive the asymptotic distribution of \( \hat{\lambda}_1 \).

**Complete data analysis**

3(a) (5 points) In this approach, we only use the data from the patients whose levels are detectable, i.e., \( X_1, \ldots, X_m \). Then the likelihood function should be

\[ f(X_1|X_1 \geq \tau) \times \cdots \times f(X_m|X_m \geq \tau), \]

where \( f \) denotes the conditional density. Explain why the likelihood function should be like this and explicitly write out the above expression.

(ATTN: more questions on the back of this page)
3(b) (5 points) Using the above likelihood function and conditional on $X_1 \geq \tau, \ldots, X_m \geq \tau$ and $m$, find the UMVUE for $\lambda$, denoted by $\hat{\lambda}_2$, and calculate its conditional variance.

3(c) (5 points) Does the UMVUE attain the Cramer-Rao bound, conditional on $X_1 \geq \tau, \ldots, X_m \geq \tau$ and $m$?

3(d) (5 points) What is the asymptotic distribution of $\hat{\lambda}_2$? The asymptotic distribution should be unconditional.

Maximum likelihood estimation

4(a) (10 points) We aim to obtain the maximum likelihood estimator for $\lambda$, denoted by $\hat{\lambda}_3$, using all the observations from $n$ patients. Since the last $(n - m)$ patients have undetectable levels, the EM algorithm can be used to calculate $\hat{\lambda}_3$. Write out the E-step and M-step explicitly.

4(b) (5 points) What is the asymptotic distribution of $\hat{\lambda}_3$?

4(c) (5 points) Construct an asymptotic 95%-confidence interval for $\lambda$ based on $\hat{\lambda}_3$.

4(d) (5 points) What are the asymptotic relative efficiencies of $\hat{\lambda}_1$ vs $\hat{\lambda}_3$ and $\hat{\lambda}_2$ vs $\hat{\lambda}_3$?