1. A study conducted in the Maternal and Child Health Department asks each of n randomly selected mothers of infants less than 6 months of age a set of k "yes" or "no" questions. Assuming the probability of answering "yes" in each question is denoted by π, the researcher in interested in estimating the probability of "perfect care", which is defined by answering "yes" in all of the k questions, i.e., $\theta = \pi^k$.

   (a) Let X_i denote the number of questions mother i answering "yes" among those k questions. One can assume that X_i follows a binomial distribution with pdf

   $$f(x|\pi) = \left(\begin{array}{c} k \\ x \end{array} \right) \pi^x (1-\pi)^{k-x}.$$

   Assuming $X_1, \ldots, X_n$ is a random sample of size n, derive the maximum likelihood estimator (MLE) of θ.

   (b) Derive Cramér-Rao Lower Bound (CRLB) of the variance of any unbiased estimator of θ.

   (c) Show that $T = \sum_{i=1}^n X_i$ is a complete and sufficient statistic for π (as well as for θ), and that

   $$W = \left\{ \begin{array}{ll} 1 & \text{if } X_1 = k \\ 0 & \text{otherwise,} \end{array} \right.$$

   is an unbiased estimator of θ.

   (d) By Lehmann-Sheffe Theorem, the estimator $\phi(T) = E(W|T)$ is the uniformly minimum variance unbiased estimator (UMVUE). Show that

   $$\phi\left(\sum_{i=1}^n X_i \right) = \frac{\left(\begin{array}{c} (n-1)k \\ \sum_{i=1}^n X_i - k \end{array} \right)}{\left(\begin{array}{c} nk \\ \sum_{i=1}^n X_i \end{array} \right)}$$

   if $\sum_{i=1}^n X_i = k, k+1, \ldots, nk$, and $\phi(\sum_{i=1}^n X_i) = 0$ otherwise. Here you may use the fact that $\sum_{i=1}^n X_i$ follows a binomial distribution with parameters nk and π, denoted by $\text{Bin}(nk, \pi)$.

   (e) Without using $E(E(W|T)) = E(W)$, show that $\phi(\sum_{i=1}^n X_i)$ is indeed an unbiased estimator of θ.

(f) One may see that the variance of $\phi(\sum_{i=1}^{n} X_i)$ is quite complicated and difficult to derive, which makes the estimator less useful because one cannot make any inference on θ (e.g., does a mother with private insurance more likely have a "prefect care"?) A biostatistician suggests using the MLE in (a) to make inference on θ since the large sample property of the estimator can be easily derived. Comment on the approach of the biostatistician and derive the large sample distribution of the MLE of θ in (a). [Hint: Use Central Limit Theorem (CLT) and Delta method].

2. In malaria study, human-to-mosquito transmission is mediated by sexual stage parasites called gametocytes. *P. falciparum* gametocytes often represent only about 1% of the total parasite load in blood, which makes the detection of the gametocytes difficult. Suppose the amount of the gametocytes detected by a current technology, e.g., Pfs25 RT-PCR, distribute like a location-shifted exponential distribution with pdf

$$f(x|\theta) = \frac{1}{\beta} \exp\left(-\frac{x-\theta}{\beta}\right),$$

where $x > \theta$ and $\theta > 0$ is so-called detection limit. Assuming β is *known*, a researcher is eager to know how low is the amount of gametocytes the current technology can detect.

(a) If the research collects a random sample $X_1, \ldots, X_n$ of size n from an essay, derive the maximum likelihood estimator $\hat{\theta}$ of θ.

(b) To test a null hypothesis $H_0 : \theta \leq \theta_0$ versus the alternative hypothesis $H_0 : \theta \geq \theta_0$, one can derive a likelihood ratio test (LRT) statistic $\lambda(\boldsymbol{x})$ and claim that the null hypothesis is rejected if $\lambda(\boldsymbol{x}) \leq c$ for some cutoff c. Derive the LRT statistic $\lambda(\boldsymbol{x})$.

(c) One may find the cutoff c is difficult to find since one may not know the distribution of $\lambda(\boldsymbol{X})$. Instead, we can find an equivalent rejection region using the MLE in (a). Show that the equivalent region is $\{\boldsymbol{x}; \hat{\theta} \geq c^*\}$.

(d) Find the cutoff c^* such that the test has a size α, i.e., $\alpha = \sup_{\theta \in \Theta_0} P(\hat{\theta} \geq c^*)$.