PROBLEM CORNER

Md S. Warasi
Department of Mathematics and Statistics
Radford University, VA 24142
email: msarker@radford.edu

Problem 1

To screen blood donors for HIV, the American Red Cross often implements pool testing, where pools are formed by compositing a set of individual donations and then the pooled samples are tested for the presence or absence of HIV; see Figure 1. A pool is positive when at least one individual in the pool has disease; however, a pool is negative when all individuals in the pool are free of disease. Unfortunately, the assay being used for diagnosis is subject to errors. When a positive pool is tested, there is a 97% probability that the test result is positive (a correct result). When a negative pool is tested, there is a 98% probability that the test result is negative (also a correct result). Assume that the individuals are independent and have an identical probability of 1% to be HIV positive. Also, assume that the test accuracy does not depend on the pool size. Suppose a pool comprised of 3 individuals is tested for HIV.

a. What is the probability that the pool tests positive?

b. Write an algorithm to approximate the probability in 1(a) by simulation.

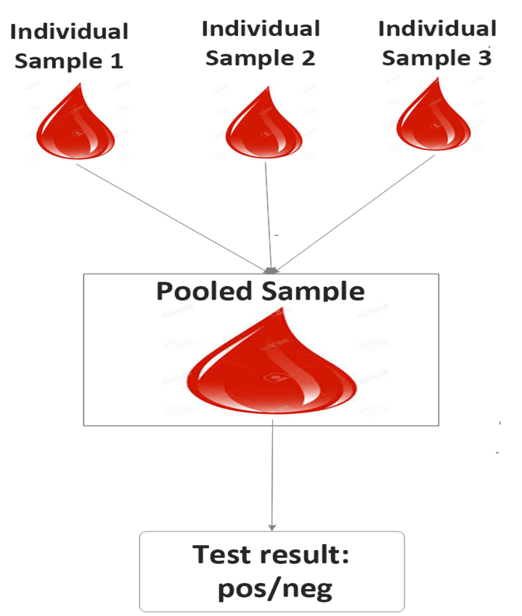

Figure 1: Pool testing to screen blood donors for HIV.

Problem 2

In statistics, maximum likelihood is a procedure of estimating the parameters of a probabilistic model. In the context of pool testing, the maximum likelihood technique is used to estimate individual-level disease prevalence using data observed from pools; see Problem 1 for more details about pool testing.

Consider a pool testing application, where μ denotes the probability that an individual has HIV. Suppose J pools, each of which is comprised of n individuals, are tested for HIV. Let z_j, for $j = 1, 2, ..., J$, denote testing responses, where $z_j = 1$ if a pool tests positive and $z_j = 0$ if otherwise. Finding maximum likelihood estimate of the parameter μ involves maximizing $L(\mu) = \prod_{j=1}^{J} \theta^{z_j}(1-\theta)^{1-z_j}$ as a function of μ, where $\theta = 1 - (1-\mu)^n$ and $\mu \in (0,1)$; i.e., if $\widehat{\mu}$ denotes the maximum likelihood estimate of μ, then $\widehat{\mu} = \arg\max_{\mu} L(\mu)$. Show that

$$\widehat{\mu} = 1 - \left(1 - \frac{\sum_{i=1}^{J} z_i}{J}\right)^{1/n}$$

and find $\widehat{\mu}$ for the following data, where $J = 10$ and $n = 4$.

Pool testing data										
z	1	0	1	0	0	1	1	1	1	1