BIOSTAT/STAT 570: Midterm Take-Home Exam

To be submitted to the course canvas site by 11:59pm Saturday 12th November, 2022. This is an exam, so no collaboration.

In this exam you will investigate different approaches to analyzing a simple prevalence estimation problem, that we looked at briefly in class.

In early April, 2020, 3330 residents of Santa Clara County, California were recruited then and tested for COVID-19 antibodies and 50 people tested positive (Bendavid et al. 2020, MedRxiv).

Suppose a seroprevalence test is carried out with sensitivity $\delta = \Pr(\text{ +ve test } \mid \text{ disease })$ and specificity, $\gamma = \Pr(\text{ -ve test } \mid \text{ no disease })$, where $0 < \gamma \leq 1$, $0 < \delta \leq 1$ and we assume $\gamma + \delta > 1$. Let π denote the true prevalence.

We test N people and y are recorded as having the disease, and a starting model is

$$y|p \sim \text{Binomial}(N, p)$$

where p is the probability of a +ve test result, with

$$\begin{aligned} p &= \Pr(\text{ +ve test }) \\ &= \Pr(\text{ +ve test } \mid \text{ disease })\Pr(\text{ disease }) + \Pr(\text{ +ve test } \mid \text{ no disease })\Pr(\text{ no disease }) \\ &= \pi(\delta + \gamma - 1) + (1 - \gamma) \end{aligned}$$

Suppose, initially, that the sensitivity and specificity are known and we want to estimate π. Assume the sensitivity is 0.8 and the specificity is 0.995.

1. **Likelihood Analysis**

   (a) **4 Points** Show that under the binomial model the MLE is

   $$\widehat{\pi} = \frac{y - N(1-\gamma)}{N(\delta + \gamma - 1)}.$$

   (b) **4 Points** Give an expression for the expected information and hence give the form for a 90% Wald confidence interval.

   (c) **4 Points** Obtain the MLE and 90% confidence interval for π the Santa Clara data.

   (d) **2 Points** Obtain the MLE and the expected information for $\theta = \log(\pi/(1-\pi))$.

(e) **2 Points** Suppose one believes the data are overdispersed relative to a binomial. Is it possible to fit an overdispersed model for the data here?

2. **Bayesian Analysis**

   We will now carry out a Bayesian analysis with a Beta(1,1) prior for π. For the methods listed below, evaluate the normalizing constant, the posterior mean and the posterior variance for $\theta = \log(\pi/(1-\pi))$. Give your results in the form of a table. Also give Monte Carlo confidence intervals on your estimates, for those methods that you can do this for.

   (a) **4 Points** Write down the functional form for the posterior $p(\theta|y)$, up to the normalizing constant.

   (b) **4 Points** Gauss-Hermite quadrature using 4 point and 5 point rules. You should decide how to center and scale the integration design points.

   (c) **4 Points** A rejection algorithm, using the prior as proposal. In addition to the normalizing constant, the posterior mean and the posterior variance, give a histogram representation of the posterior distribution for π.

   (d) **4 Points** Importance sampling Monte Carlo. Use a t-distribution with 4 degrees of freedom, with mean and scale based on the MLE and its asymptotic variance.

   (e) **4 Points** A Metropolis algorithm based on a normal proposal distribution centered at the current point with a variance you should choose. In addition to the posterior mean and the posterior variance, give a histogram representation of the posterior distribution for π (For this part, you do not need to evaluate the normalizing constant).

   (f) **6 Points** Suppose that rather than being taken as fixed, beliefs about the sensitivity δ and the specificity γ can be represented via beta distributions. Specifically, $\delta \sim$ Beta$(160, 40)$ and $\gamma \sim$ Beta$(1990, 10)$. Extend your approach in (b) or (d) to obtain samples from the posterior for π, based on these new beliefs about δ and γ. How does the posterior change relative to the point estimates for δ, γ analysis?

   (g) **BONUS 4 Points** Laplace approximation.

Hint: Section 3.7 of Wakefield (2013) contains many of the details you will need for this work, and the book website contains `R` code to carry out similar analyses.