Math 152 - Statistical Theory - Homework 10

write your name here

Due: 12/7/2018

Suppose that for a particular tropical disease no gold standard clinical test exists. Instead we have a test that is known to be imperfect; not always identifying a disease if the patient has the disease, and sometimes yielding false positives (patients that do not have the disease but test positive). However, by using this test in a clinical trial it is hoped that we can obtain better estimates for the disease sensitivity (S; the proportion of disease positive individuals who test positive) and specificity (C; the proportion of individuals who don't have the disease who test negative).

To do this we can construct a table of the observed and latent data for the test outcomes. In the table a and b are the number of observed positive and negative results respectively. Y1 and Y2 are latent variables that represent the gold standard – the true number of positive individuals out of a and b respectively.

		Truth		
		+	-	
Test	+	Y_1	$a - Y_1$	a
	-	Y_2	$b - Y_2$	b
		$Y_1 + Y_2$	$N - Y_1 - Y_2$	N

1. Write down an expression for the likelihood, supposing that the prevalence for the disease is π. Hint: multiply together the likelihoods corresponding to each of the interior cells in the table.
2. Assuming priors of the form: $\pi \sim beta(\alpha_\pi, \beta_\pi)$, $S \sim beta(\alpha_S, \beta_S)$ and $C \sim beta(\alpha_C, \beta_C)$, show (derive) that the marginal conditional probabilities are:

$$
\begin{aligned}
Y_1|a,\pi,S,C &\sim binomial\left(a, \frac{\pi S}{\pi S + (1-\pi)(1-C)}\right) \\
Y_2|b,\pi,S,C &\sim binomial\left(b, \frac{\pi(1-S)}{\pi(1-S) + (1-\pi)C}\right) \\
\pi|a,b,Y_1,Y_2 &\sim beta(Y_1 + Y_2 + \alpha_\pi, a + b - Y_1 - Y_2 + \beta_\pi) \\
S|Y_1,Y_2 &\sim beta(Y_1 + \alpha_S, Y_2 + \beta_S) \\
C|a,b,Y_1,Y_2 &\sim beta(b - Y_2 + \alpha_C, a - Y_1 + \beta_C)
\end{aligned}
$$

Hint: for the binomial start with the marginal conditional and show that the kernel of the distribution is the same that you found in the joint distribution.

3. Suppose that out of a sample of 100 people, 20 of those tested negative and 80 positive. Assuming uniform priors on π, S and C, use Gibbs sampling to generate posterior samples for π. What do you conclude?

Note: uniform priors are the same as Beta(1,1)

4. Suppose that a previous study that compare the clinical test with a laboratory gold standard concludes that $S \sim beta(10, 1)$ and $C \sim beta(10, 1)$. Use Gibbs sampling to estimate the new posterior for π. Why does this look different to your previously-estimated distribution?

5. Suppose a previous analysis concluded that $\pi \sim beta(1, 10)$. Using this distribution as a prior, together with uniform priors on S and C, determine the posterior distributions for the test sensitivity and specificity respectively. Why does the test appear to be quite specific, although it is unclear how sensitive it is?
6. Suppose that based on lab results you suppose that the test specificity $C \sim beta(10, 1)$, and $\pi \sim beta(1, 10)$, but the prior for S is still uniform. Explain the shape of the posterior for S now.
7. Now suppose that the sample size was 1000 people of which 200 tested positive. Using the same priors as the previous question, determine the posterior for S. What do you conclude about your test's sensitivity?
8. What do the previous results (across the entire assignment) suggest is necessary to assess the sensitivity of a clinical test for a disease?

[This problem comes from "A Student's Guide to Bayesian Statistics" by Ben Lambert]