# BAYESIAN EXERCISES

(1) We have available a test for whether a person has a rare disease. The test is always either positive (T=1) or negative (T=0). If the person is sick (S=1), the test is positive with probability .99, and if the person is well the test is negative with probability .99.

(a) Show that the random subset of possible $S$ values that is $\{1\}$ when $T = 1$ and $\{0\}$ when $T = 0$ is a 99% confidence set for the unknown $S$ value.

(b) Suppose we are administering this test to patients drawn randomly from a population in which we know the rate of occurrence of this disease is one in 10,000. What is the conditional probability that a single patient drawn from this population has the disease, given that the test is positive?

(c) Suppose instead that we have been flown in with our testing equipment to a remote town where the prevalence of the disease is unknown. Suppose we take our prior beliefs about the probability $p$ of a randomly drawn individual from this town being sick as a uniform distribution on $(0, 1)$. Now suppose that we test 100 randomly drawn people in the town and find one positive test, with the other 99 negative. Given this evidence and the uniform prior, what is our posterior probability that the person with $T = 1$ actually has $S = 1$? What is our posterior probability that none of the other 99 people with $T = 0$ have the disease?

(2) This is an example where coming up with a nicely behaved confidence interval is a challenge, while Bayesian credibility sets are fairly straightforward.

Suppose $\{X_i, \dots, X_N\}$ are i.i.d. random variables, each with the distribution $N(\mu, \mu^2)$, where $\mu$ is an unknown parameter lying in $(-\infty, \infty)$. As with any normally distributed random variables, sufficient statistics for the unknown parameter $\mu$ are the first and second moments $m_1 = \sum X_i/N$ and $m_2 = \sum X_i^2/N$. There are several ways to construct confidence intervals for $\mu$ based on **pivotal quantities**, that is functions of $\mu$ and the data that have a distribution that does not depend on $\mu$. Three such pivotal quantities are

$$\frac{m_1}{\mu} \sim N(1, 1/N)\,, \qquad \frac{N(m_2 - m_1^2)}{\mu^2} \sim \chi^2(N-1)\,, \qquad \sum_i \left(\frac{X_i}{\mu} - 1\right)^2 \sim \chi^2(N)\,. \tag{1}$$

There are two standard ways to construct Bayesian credibility sets: highest-posterior-density, or HPD sets, and equal-tail intervals. The equal-tail intervals are easiest to construct: for a probability $1 - \alpha$ equal tail credibility set, compute the posterior density and find the points $a, b$ that leave probability $\alpha/2$ below $a$ and above $b$. If you need to do this numerically (as you probably do in this problem) you compute the density at a large number of points, normalize it to sum to one, and then find the $a, b$ pair. An HPD set, which may not be a single interval, collects all parameter

*Date*: February 6, 2020.

values for which the posterior density exceeds some level $\gamma$, and chooses $\gamma$ so that the probability of the set is $1-\alpha$. This approach gives the credibility set of smallest possible total length. It requires sorting the $\mu$ values by their associated posterior density values, then finding the cutoff $\gamma$ that gives the desired probability $1-\alpha$.

(a) Show that when the sample size is $N=1$, only one usable pivotal quantity emerges from the three formulas above. Also show that the likelihood for $N=1$ is not integrable.

(b) Give formulas for confidence sets based on the three pivotal quantities. There is not a unique way to do this. You might use one or two-tailed intervals with the $\chi^2$ distributions, for example. But you need only do one for each pivotal quantity. Be sure to recognize that the mapping from an interval for the pivotal quantity to an interval for $\mu$ will involve manipulating inequalities, which may require distinguishing cases where $\mu$ is negative from those where it is positive, for example. Also note that the set may not be a single interval and may not be bounded.

(c) For each of the following hypothetical samples, calculate the maximum likelihood estimator of $\mu$, the posterior mean of $\mu$ under a flat prior, 95% confidence intervals from at least two of the three pivotal quantities and 95% credibility sets, both HPD and equal-tailed, for $\mu$. In some cases this will be impossible (e.g. in the sample with $N=1$), in which case you can just point that out. Display the intervals on a plot of the likelihood function. The samples:

$$\{1\} \tag{2}$$
$$\{1,-1\} \tag{3}$$
$$\{2, 2.1, 2.05\} \tag{4}$$
$$\{-.2, .3, .5\}\ . \tag{5}$$

(d) Comment on whether there seems to be a sense in which some of the confidence sets are better behaved than others. Also on which are most similar to Bayesian credibility sets.

(3) **Locating a break date.** Suppose we have observations on $X_t$, $t=1,\dots,T$ and that for $t \le T^*$, the $X_t$'s are i.i.d. $N(\mu_1, \sigma^2)$ and for $t > T^*$ they are distributed as $N(\mu_2, \sigma^2)$.

(a) If $\sigma^2$ is known and we use independent flat priors on $T^*$, $\mu_1$ and $\mu_2$, show how to form the marginal posterior distribution of $T^*$ and an HPD credible set for it.

(b) Think about how one might try to construct a confidence interval for $T^*$. (This is not at all easy.)

(c) What if $\sigma^2$ is unknown?

(d) If variances change also at $T^*$, so the distributions are $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, flat priors on $\sigma_j^2$ and $\mu_j$ no longer give reasonable results. Why?