Bayesian Statistics
Exercises 1.

Note: the Gamma(m, λ) distribution refers to the distribution with density $p(x) \propto x^{m-1} \exp(-\lambda x)$

1. Suppose $x \in \{1, 2, 3\}$ and $\theta \in \{0, 1\}$ with

x	1	2	3
$p(x\vert\theta = 0)$	0.005	0.005	0.99
$p(x\vert\theta = 1)$	0.0049	0.9851	0.01

   Now consider testing the null hypothesis, $H_0 : \theta = 0$, vs $H_1 : \theta = 1$ using a likelihood ratio test. Recall that a type I error is defined as rejecting H_0 when it is true, and a type II error is failing to reject H_0 when it is false.

   (a) Consider the test that rejects H_0 when $x \in \{1, 2\}$. Derive the probability of a type I error given that H_0 holds, and the probability of a type II error given that H_0 does not hold.

   (b) If $x = 1$, what is the p value? What is the Bayes Factor for H_1 vs H_0? Comment on how strong (qualitatively) the evidence is to reject H_0.

   (c) If $x = 2$, what is the p value? What is the Bayes Factor for H_1 vs H_0? Comment on how strong (qualitatively) the evidence is to reject H_0.

   (d) Modify the example so that $x = 1$ corresponds to a p value of 0.01 and a very large Bayes Factor in favour of H_1.

2. Sellke et al. (2001) consider the problem of calibrating p values in testing a series of drugs. In particular, they consider testing a series of drugs, and ask "of the tests that produce a p value near 0.05, what proportion of the corresponding drugs follow the null?". As discussed in class, the answer to this question will depend on i) the overall proportion of drugs for which the null hypotheses is true (π), and ii) the effect size of the non-null drugs (θ_1). [It will also depend on the sample

size used to do the test; you may choose to fix this or allow it to vary, as you prefer.]

Perform a simulation study (e.g. using R), to illustrate the way that the answer depends on i) and ii) above. Your answer should have the following structure: a) A brief, precise, and self-contained description of the situation you considered (this should be at a level that it would make sense to another graduate student who has not read Sellke et al); b) a summary of the results you obtained, in table or graph form as you feel appropriate; c) a discussion of how your results compare with those in Sellke et al.. Include any computer code you used as an appendix.

3. Consider $x_1, x_2, \ldots, x_n$ i.i.d. with sampling distribution $f(x|\mu)$ which is Poisson(μ). Suppose the prior on μ is Gamma(m, λ). Find the posterior for μ.

4. Suppose that, given σ, x is Normal with mean 0 and variance σ^2, and that the prior for σ is such that $1/\sigma^2$ has a Gamma(1,2) distribution. Find the posterior distribution of σ^2. [Hint: the algebra is easier if you compute the posterior for the "precision", $\tau := 1/\sigma^2$, rather than for the variance σ^2. For this reason Bayesian analyses often consider the precision rather than the variance.]

5. Clone or fork the repository `https://github.com/stephens999/stat302/`. Run the R code in `exercises/seeb/train_test.R`, which loads and processes a dataset on 544 fish (salmon) at 12 genetic markers/loci, from Seeb et al. (2007). Complete the exercise in the commented text at the end of `train_test.R`.

References

Seeb, L., A. Antonovich, M. A. Banks, T. Beacham, M. Bellinger, S. Blankenship, M. Campbell, N. Decovich, J. Garza, C. Guthrie III, et al. (2007). Development of a standardized DNA database for Chinook salmon. *Fisheries 32*(11), 540–552.

Sellke, T., M. Bayarri, and J. Berger (2001). Calibration of p values for testing precise null hypotheses. *The American Statistician 55*(1), 62–71.