Bayesian Analysis I, Spring term 2020, TAKE-HOME Assignment (due date: 20.05)

INSTRUCTIONS: You may use any books, references, notes and computers. but are not allowed to discuss this home-task with any person other than the instructor until the due date. Be sure to show **all** of your work **in detail**. Please put your *name* and *student no.* in your writeup and SUBMIT all the computer codes to the moodle. You can also scan any handwritten results and upload them to the moodle.

1. As part of a research project, geologists are interested in rock which degrades at a rate of more than 0.12mm per year. Gravestones provide a useful means of measuring this rate. Let θ be the probability that, in a particular cemetery, a gravestone indicates an annual rate of degradation of more than 0.12mm.

Of the twenty gravestones in a randomly chosen cemetery, twelve show an annual rate of degradation of more than 0.12mm.

a) Using a noninformative prior, Unif(0,1), obtain posterior inference for θ. Also, make the overlaid plots of prior and posterior distributions.

b) Using a specific Beta(10,2) prior, obtain posterior inference for θ. Make the overlaid plots of prior and posterior distributions.

c) Describe briefly your findings regarding two different priors above. Draw a short conclusion on this analysis for θ.

2. A cancer laboratory is estimating the rate of tumorigenesis in two strains of mice, A, and B. They have tumor count data for 10 mice in strain A and 13 mice in strain B. Type A mice have been well studied, and information from other laboratories suggests that each type A mouse has a tumor count that is approximately Poisson-distributed with a mean of 12. Tumor count rates for type B mice are unknown, but type B mice are related to type A mice.

The observed tumor counts for the two populations are

A	12 9 12 14 13 13 15 8 15 6
B	11 11 10 9 9 8 7 10 6 8 8 9 7

a) Find the posterior distribution, mean, variance and a 95% HPI (credible) interval for θ_A, assuming a Poisson sampling model and the prior distribution $\text{Gamma}(120, 1/10)$.

b) Find the posterior distribution, mean, variance and a 95% HPI (credible) interval for θ_B, assuming the same Poisson model and the prior distribution $\text{Gamma}(12, 1)$.

c) Make the plots of prior and posterior distributions (overlaid) for θ_A and θ_B, respectively.

d) Compute the posterior expectation of θ_B under the prior distribution $\theta_B \sim \text{Gamma}(12n_0, n_0)$ for each value of $n_0 \in \{1, 2, ..., 50\}$. Describe what sort of prior beliefs would be necessary in order for the posterior expectation of θ_B to be close to that of to be θ_A.

3. Eight college statistics majors are playing basketball together for the first time. Just for the fun of it, they decide to use Bayesian methods to estimate each individual student's

success probability of making a basket from 10 feet away, as well as the overall average success probability of the group. They gather data in the following way:

One at a time, each student stands at a position 10 feet from the basket and keeps shooting until he finally makes a basket. Another student records x_i -how many failures shooter i had before successfully making his first basket.

The geometric probability mass function is appropriate for modeling each student's number of failures before the first success. The students have never played together before, so they don't have any knowledge about who the good shooters and poor shooters might be. Thus, in their Bayesian model, they consider the success probabilities $p_i, i = 1, ..., 8$ to be random draws from a common Beta density. They complete their Bayesian model by specifying priors on the parameters of the Beta density.

Note that the geometric distribution is a special case of the negative binomial distribution, namely negative binomial with the second parameter equal to 1. In BUGS,

$x \sim \text{dnegbin}(p, 1)$

says that x is drawn from a geometric distribution with the parameter p.

Each student's number of failures before the first goal were 4, 0, 1, 7, 3, 2, 8, 0.

a) Using a Gamma(1,1) hyper-prior for both of the parameters (α, β) of the beta prior, find the posterior distribution, mean, variance and the 95% credible interval for p_i's. (You can do this using either WinBUGS or RJAGS.)

b) If the students had used a Gamma(1,2) on the parameter β instead of Gamma(1,1), would that have been likely to make any difference in the resulting posterior means of individual p_i's? Explain why or why not briefly.

(Recall that in BUGS, the gamma distribution is specified with $1/\beta$ in our list of distributions.)

c) Obtain the posterior inference for $\theta = \frac{\alpha}{\alpha+\beta}$.

d) By choosing different initial values for α and β, explore the sensitivity of your analysis. Briefly report your findings and give some conclusions on your analysis.

4. A study was conducted on 32 cars to explore the relationship of the gasoline consumption on the the weight of the car and engine sizes in cylinders.

For each car we have observations on how many miles that car can travel on a gallon of gasoline (mpg), the weight of the car (weight) and two dummy variables that indicates if the car's engine has four cylinders (sixcyl=0 and eightcyl=0) six cylinders (sixcyl=1 and eightcyl=0) or eigth cylinders (sixcyl=0 and eightcyl=1).

mpg: 21.0,21.0,22.8,21.4,18.7,18.1, 14.3, 24.4, 22.8,19.2,17.8,16.4,17.3,15.2,10.4,10.4,14.7, 32.4,30.4,33.9, 21.5, 15.5,15.2,13.3,19.2,27.3, 26.0,30.4,15.8,19.7,15.0, 21.4

weight: 2.620, 2.875, 2.320, 3.215, 3.440, 3.460, 3.570, 3.190, 3.150,3.440, 3.440, 4.070, 3.730, 3.780, 5.250, 5.424, 5.345, 2.200,1.615, 1.835, 2.465, 3.520,3.435, 3.840, 3.845, 1.935, 2.140,1.513, 3.170,2.770, 3.570, 2.780

sixcyl: 1, 1, 0, 1, 0, 1, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0

eightcyl: 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 1, 1, 1, 1, 0, 0, 0, 1, 0, 1, 0

We want to sample from the joint posterior distribution in the Gaussian linear regression

$$mpg = \beta_0 + \beta_1 * weight + \beta_2 * sixcyl + \beta_3 * eightcyl + error, \quad error \sim N(0, \sigma^2)$$

with conjugate priors

$$\begin{aligned} \beta_i &\sim N(0, 10000), \quad i = 0, ..., 3 \\ 1/\sigma^2 &\sim Gamma(0.01, 0.01) \end{aligned}$$

a) Give the plots of the marginal distributions for each parameter.

b) Construct 95% equal tail probability intervals for each parameter and interpret couple of them.

c) Investigate if the effect on mpg is different in cars with six cylinders compared to cars with 8 cylinders.

d) Estimate the predictive distribution for a new 4 cylinder car with weight = 3.5.

5. A new process is proposed for preparing chicken breasts. One measure of its success will be the level of wastage; the standard process produces an average 4.7 % wastage. The process is tested by two operators on 6 chicken breasts each. The data were as follows (in % wastage):

Operator 1	4.3, 4.3, 2.7, 3.6, 3.5, 4.5
Operator 2	3.9, 4.0, 4.5, 2.9, 5.2, 4.8

It is thought that the mean μ should be the same for both operators, but that they may have different variances. If y_{ij} is the wastage for observation j from operator i, the model is that

$$y_{i,j}|\mu, \tau_1, \tau_2 \sim N(\mu, \tau_i)$$

independent. The parameters μ, τ_1, and τ_2 have independent prior distributions

$$\mu \sim N(4.7, \nu), \quad \tau_i \sim IG(4, 5)$$

a) The elicited value of ν is 0.2. Use WinBUGS or RJAGS to analyze these data for the case $\nu = 0.2$, with particular reference to inference about whether $\mu < 4.7$ and about the posterior distribution of $\phi = \frac{\tau_2}{\tau_1}$. [Your solution should include a listing of your WinBUGS/RJAGS program and output.]

b) Explore sensitivity of the analysis to ν, considering the range $0.15 \leq \nu < \infty$. i.e. select several values of ν in the given range and check how the results change according to them.

c) Write a short report on your findings for the process developer who is assumed not to be statistician. i.e. Explain your conclusions in simple language for the developer.