

# Exam Bayesian Methods

The exam will be available on 6 July 2017, 8:00.

Solutions will be submitted (sent via email to [oliver@kirchkamp.de](mailto:oliver@kirchkamp.de) before 6 July 2017, 9:00. A late submission penalty applies.

For all your answers write down (all) the commands you use to obtain these answers. Also include the answers in a short form (for each question do not include more than one line of output – usually your answer should just be one or two numbers).

1. A random variable  $X$  follows a normal distribution with known mean  $\mu$  and unknown precision  $\tau$ . Your prior for  $\tau$  follows a Gamma distribution with mean 2 and variance 1.

You observe a sample  $x$  of size  $n = 10$ , mean  $\mu_x = 10$  and variance  $\sigma_x^2 = 4$ . What is the posterior distribution of  $\tau$ ?

Hint: If  $X \sim \Gamma(\alpha, \beta)$  then  $E(X) = \frac{\alpha}{\beta}$  and  $\text{var}(X) = \frac{\alpha}{\beta^2}$ .

2. Consider the Crime dataset from Ecdat. Use only the year 1981 and explain in a linear model `crm rte` as a function of `prbarr`, `prbconv`, `prbpris`, `avgsen`, `polpc`, `density`, `taxpc`, `pctmin`, `wcon`, `wtuc`, `wtrd`, `wfir`, `wser`, `wmfg`, `wfed`, `wsta`, `wloc`, `mix`, `pctymle`. (i.e.: all variables except `county`, `year`, `crm rte`, `region`, `smsa`).

Use a vague prior for the regression coefficients.

- a) Use `jags` to estimate the posterior distribution. First, do not demean your variables. (10000 iterations take 4 seconds on an i7-6700K).
  - b) What can you say about the potential scale reduction factor? What can you say about effective sample size?
  - c) Give a credible interval for the standard deviation of your residuals.
  - d) Give a credible interval for the coefficient of `polpc`.
  - e) Now demean your variables. What changes do you observe?
  - f) Determine a credible interval for the difference between the coefficients of `prbarr` and `prbconv`.
  - g) Determine a credible interval for the ratio of the coefficients of `prbarr` and `prbconv`.
  - h) Consider the hypothesis that the coefficient of `polpc` is larger than 2.5. Based on your posterior, what are the odds that this statement is true?
  - i) Consider the hypothesis that the ratio of the coefficient of `prbarr` and the coefficient of `prbconv` is larger than 1.5. Based on your posterior, what are the odds that this statement is true?
3. Pretend that (for the same dataset) you can only observe whether `crm rte` is larger than .03 or whether it is not larger than .03. Estimate a probit model where this binary

variable is the dependent variable. Everything else remains as above (work with the demeaned data, but do not demean your dependent (binary) variable!).

(10000 iterations take 55 seconds on an i7-6700K).

- a) Give a credible interval for the coefficient of `polpc`.
  - b) Determine a credible interval for the difference between the coefficients of `prbarr` and `prbconv`.
  - c) Determine a credible interval for the ratio of the coefficients of `prbarr` and `prbconv`.
  - d) Consider the hypothesis that the coefficient of `polpc` is larger than 2.5. Based on your posterior, what are the odds that this statement is true?
  - e) Consider the hypothesis that the ratio of the coefficient of `prbarr` and the coefficient of `prbconv` is larger than 1.5. Based on your posterior, what are the odds that this statement is true?
4. Now include in your previous model (the one where your dependent variable is binary) a random effect for `region`.

(10000 iterations take 66 seconds on an i7-6700K).

- a) What is the standard deviation of the random effect?
- b) Give a credible interval for the coefficient of `prbconv`.
- c) What is now a credible interval for the difference between the coefficients of `prbarr` and `prbconv`.