This exercise sheet constitutes the exam homework, which needs to be handed in separately by each student signed-up for the course in order to obtain the ECTS credits for the course. Your solution shall consist of the following:

- A written report as a PDF file containing solutions in the form of results, textual interpretations and graphs for the four homework exercises. Note: plagiarism or other forms of cheating is a serious act – to underline this your report must as cover page contain the signed confirmation that your work is made in accordance with the Rules for Written Exams at Stockholm University. For further information about possible consequences see also the Regulations for Disciplinary Matters at Stockholm University.
- A file `<lastname>.R` containing the R code used to obtain all results and graphics contained in the report. Structured and well-documented code is important, e.g., each function should be preceded by a short text explaining what the input and output parameters are. Further code comments are to be made where needed and indentation should be used – see, e.g., Google's R Style Guide for further guidelines. Results are not to be discussed in the code – this is done in the report. As a trivial quality check: the command `source("<lastname>.R")` should run without errors for your code file.
- Deadline: **Wednesday 03 Nov 2019 at 18:00 o'clock**. The report has to be handed in as a bundle consisting of a) A scanned copy of your signed `Confirmation.pdf`, b) a PDF file `<lastname>.pdf` containing your report, c) the R file `<lastname.R>` and d) (in case of Sweave/knitr) `lastname.R[nw|md]` before the deadline. If you modified the original data or if your R code relies on external files, your bundle should contain these files as well (optimally as a ZIP file). All files are to be uploaded before the deadline to the Moodle drop-box on the course home page. Please note that there is a 10Mb file limit when uploading files. Delayed hand-ins are not accepted.

A total of 100 points can be reached for the answers in the report. Furthermore, up to 5 additional bonus points can be obtained, should your report and code be written with knitr. In this case please also attach the file `<lastname>.R[nw|md]` to your upload. Your final grade is determined by your sum of regular points and bonus points. Note: A penalty is imposed on reports longer than 30 pages.

Lycka till!

Exercise 1 (22 points)

This exercise contains some of the recommended exercises given in the class. Clear steps must be shown to receive full points. For the next 3 questions, suppose the likelihood $f(x|\theta,\sigma)$ is normal with mean θ and standard deviation σ:

1. (2 points) Suppose σ is known, find the Jeffrey's prior for θ.
2. (2 points) Suppose θ is known, find the Jeffrey's prior for σ.
3. (4 points) Suppose both θ and σ are unknown, find the bivariate Jeffrey's prior for (θ,σ). Discuss the difference from the two previous results.
4. (6 points) In class, we have shown for the univariate case that the Jeffrey's prior $p(\theta) \propto [I(\theta)]^{1/2}$ is invariant to 1-1 transformation. Show that for the multivariate case, the Jeffrey's prior $p(\theta) \propto |\mathbf{I}(\theta)|^{1/2}$ is invariant under parameter transformation, where $|\cdot|$ denotes the determinant, and $\mathbf{I}(\theta)$ is the expected Fisher information matrix. Notation of this part is the same as those in the course book.
5. (8 points) From the paper by Tierney and Kadane 1986 on Laplace approximations, derive Eq. (A.1) by showing that the leading correction term has the form a/n with the coefficient a as shown in the paper, where n is the data size. Note: you do not need to work out the higher order terms, i.e., b/n^2 and $O(n^{-3})$.

Exercise 2 (28 points)

After completing all classes from the course, you are now ready to explore an emerging field in modern Bayesian methods termed "Approximate Bayesian Computation (ABC)". To complete this exercise, you have to read the review article by Sunnaker *et al.*, "Approximate Bayesian Computation", PLOS Comput. Biol., 9:e1002803 (2013), available in `https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1002803`. Although the article discussed applications in biological science, pre-knowledge of biology is not needed in understanding the methods. *IMPORTANT*: The answers of this exercise must be written in your own words. Moreover, a clear, concise and logical writing is required to obtain full points. Feel free to use bullet points to organize your writing. Please also note that there are limitations in word counts in this exercise, writings that exceed significantly the word limits will result in point deduction!

1. (3 points) Summarize in what situations and scenarios ABC are useful. Word limit: $<$ 100 words
2. (3 points) Derive the first equation in the section"Bayes Factor with ABC and Summary Statistics".
3. (6 points) A nice conceptual overview (Fig. 1) was provided in the review concerning parameter estimation in ABC. Draw a similar conceptual overview for model comparison using ABC (refer to the section "Model Comparison with ABC").
4. (10 points) Many possible risks, remedies, extensions and generalizations of ABC have been discussed in the review. Please choose one that interests you the most, then pick up ONE corresponding reference either from the citations in the review or from the recent literatures of ABC. First specify the reference you pick, then discuss and elaborate concisely on the aspect of ABC you are interested in without the need to fully read this additional reference in details. E.g. one may discuss and elaborate how dimensionality reduction techniques can be used to deal with the curse-of-dimensionality. You can also think of this part of exercise as a practice in writing a mini-study-plan for a master thesis. Word limit: $<$ 300 words
5. (3 points) Please write down you own thoughts, perspectives, suggestions and criticisms about ABC. Word limit: $<$ 150 words
6. (3 points) List up 3 concepts mentioned in the review that you are not familiar with and would like to learn more. For each of the 3 concepts, specify a potential reference where you can learn about the concept and briefly explain why the reference is picked. Word limit: $<$ 120 words

Exercise 3 (35 points)

In this exercise you have to read the publication by Raats and Moors (2003) (available from the course home page) and re-implement some of their analyses using a) JAGS and b) your own Metropolis MCMC sampler.

(a) (6 points) Write a 1/2 - 1 page summary explaining what were the questions of interest in the paper and how were these questions were answered using a Bayesian model. Note: Use mathematical notation to describe the model – if possible even as a Bayesian graphical model in order to represent the conditional independence structure of the variables in the model. Hint: For a concise description of Bayesian graphical models see, e.g., Chapter 8 of Bishop (2006).

(b) (3 Points) Derive an expression for the likelihood $f(\boldsymbol{y}|\boldsymbol{p})$ of the model, where $\boldsymbol{p} = (p_0, p_{1|0})'$, i.e. in the situation where $p_{0|1} = 0$ and where the data are $\boldsymbol{y} = (n_1, c_0, n_2, c_{0+}, c_{01}, c_{10})'$.

Use your result together with the other assumptions used in the paper, to write up an expression, which determines the posterior distribution $p_0, p_{1|0}|\boldsymbol{y}$ up to proportionality.

Note: Opposite to the paper we will not distinguish between small and capital letters for the random variables, i.e. p_0 is used to denote both the parameter p_0 and the random variable P_0.

Hint: Please note that, to the best of our opinion, the derivations leading to eqn. (8) in the paper appear to be wrong. Instead use a formulation involving $\boldsymbol{\pi} = (\pi_0, \pi_{1|0}, \pi_{0|1})'$, i.e. based on the assumptions stated in eqn. (4) of the paper.

(c) (4 Points) Write a JAGS model implementing the model of the paper in the situation where $p_{0|1} \equiv 0$, i.e. you want to sample from the posterior $p_0, p_{1|0}|\boldsymbol{y}$. We will from now on use the non-informative prior defined by $\alpha = \beta = \epsilon = \delta = 1$. Hint: Again, use $\boldsymbol{\pi}$ as part of your model description.

(d) (4 Points) Use JAGS to generate a sample of size 10,000 from the posterior, where you as $\boldsymbol{y}$ use the data from the paper's Table 2. Plot your results for p_0 and interpret them in the context of the paper's auditing application.

(e) (6 Points) Write an R function `rdca(n, y, theta0)`, which given the previous data $\boldsymbol{y}$ implements an Metropolis sampler with component wise proposal distributions $U(0,1)$ for sampling from the posterior

$\boldsymbol{\theta}|\boldsymbol{y}$, where $\boldsymbol{\theta} = (p_0, p_{1|0})'$. Hint: Write a small helper function `dpost_unorm(theta, y)` which returns the un-normalized posterior density evaluated at $\boldsymbol{\theta}$. Note: Please show both your `dpost_unorm(theta, y)` and your `rdca(n, y, theta0)` function as part of the hand-in report.

(f) (4 Points) Use your `rdca(n, y, theta0)` function to draw three samples of size 10,000 from the posterior $p_0, p_{1|0}|\boldsymbol{y}$. Perform a convergence check of your three Markov chains and use this to perform an adequate selection of burn-in for your chains. Verify, that your MCMC implementation is correct. Describe and discuss your approach.

(g) (4 Points) Write an R function `dpost_p_0(p_0, y)`, which based on your `dpost_unorm` function uses numerical integration to compute the marginal posterior density of $p_0|\boldsymbol{y}$ evaluated at `p_0`. Hint: Use R's `integrate` function.

(h) (4 Points) Use numerical optimization to determine the posterior mode and compare your results to the paper. Furthermore, use your function to overlay a plot of the density on top of a histogram of the JAGS samples produced in part (d). Do the two densities match?

Exercise 4 (15 points)

This exercise is about performing optimal Bayesian decisions.

(a) (6 points) Let $Y_1, \ldots, Y_n|\theta \overset{\text{iid}}{\sim} \text{Po}(\theta)$ with unknown parameter $\theta > 0$. Assume that the prior distribution for θ is $\mathcal{G}a(\alpha, \beta)$ with known $\alpha > 0$ and $\beta > 0$ and expectation α/β. We want to estimate the value of θ from data, but collecting each observation y_i has a cost of c units. Thus, collecting $\boldsymbol{y} = (y_1, \ldots, y_n)'$ has a cost of $c \cdot n$ units. Consider an optimal Bayes estimator for θ in the context of the loss function specified by the equation

$$L(\theta, \hat{\theta}) = (\theta - \hat{\theta})^2.$$

Show that the optimal number of observations to use for estimating θ in this setting is given by

$$n = \sqrt{\frac{\alpha}{c\beta}} - \beta.$$

(b) (5 Points) Reconsider the model from Exercise 3 and assume as before that $p_{0|1} = 0$. Let $n_1 = 500$ and assume the following priors: $p_0 \sim Be(7, 102)$ and $p_{1|0} \sim Be(3, 4)$.

Assume that the cost of having the expert investigate a record is e=€200 and that the average loss of an incorrect record, which is misclassified as being correct by the auditor but not double checked by the expert, is a_{10}=€10,000. Derive an expression or describe an algorithm to compute the expected loss as a function of n_2.

(c) (4 Points) Plot your results from part (b) for $n_2 = 0, \ldots, n_1$ and answer the following question: What is the optimal Bayes decision for n_2 in this case? Discuss your results, in particular with respect to the process of statistical auditing.

References

Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer, New York.

Raats, V. M. and Moors, J. J. A. (2003). Double-checking auditors: a bayesian approach. Journal of the Royal Statistical Society: Series D (The Statistician), 52(3):351–365.