Practicum - Consultation: 14.04.2022, 09-11

Please upload all files to OLAT on time.

What	File name	When
Individual project	"04worksheet-Your-Name.zip"	25.04.2022 at 7 am.
Group solutions	"04worksheet-Group-Name.zip"	25.04.2022 at 7 am.
Group contribution	"05contribution-Group-Name.zip"	26.04.2022 at 22 pm.

Individual tasks
Your 04worksheet-Your-Name.zip file contains reproducible code necessary to generate your results and your report together with the resulting pdf-file.

Exercise 1 (Individual project (Part 4))

This exercise deals with an interesting approach to plan a study. At this stage there is no data but only hypotheses about the relation of the treatment an placebo groups. In the study of Baeten et al. (2013), the proof of concept (POC) requires that the ASAS20 response rate on secukinumab is larger than that on placebo. With data from 20 patients on secukinumab and 5 patients on placebo, the study should be able to show that POC $> 90\%$, for true response rates of 25% on placebo and 60% on secukinumab.

Remark: For reasons discussed in the lecture, do not use posterior distributions mentioned by Baeten et al. (2013) but base your computations on hypotheses for true response rates in placebo and secukinumab groups instead.

1. Provide code and computations in R supporting the evidence that POC $> 90\%$ for the study design, sample sizes, and response rates assumed. Provide an estimate of the Monte Carlo SE(POC).
2. Show that the 4:1 study design based on 20 patients on secukinumab and 5 patients on placebo considers the smallest number of patients for response rates (25% on placebo and 60% on secukinumab) and the condition POC $> 90\%$.
3. Explain this approach to a client.

Report your results.

Group tasks

Your 04worksheet-Group-Name.zip (one per group) file contains reproducible code necessary to generate your results and your report together with the resulting pdf-file, which can contain scans of your handwritten solutions.

Exercise 2 (Gibbs sampler)

Use the file `04GibbsSampler.R` provided on OLAT. Let $y_1, \ldots, y_n \sim \mathrm{N}(\mu, \sigma^2)$, where $\mu \sim \mathrm{N}(\mu_0, \sigma_0^2)$ and $1/\sigma^2 \sim \mathrm{G}(a_0, b_0)$.
(a) Complete the derivations from the lecture by proving that if $p(x) = \exp(-0.5(ax^2 - 2bx))$ then $X \sim \mathrm{N}(b/a, 1/a)$;
(b) Complete the following tasks:

1. Generate data: a random normal sample by assuming set.seed(44566), $n = 30$, $\mu = 4$ and $\sigma^2 = 16$.
2. Set the parameters of the prior distributions to $\mu_0 = -3$, $\sigma_0^2 = 4$, $a_0 = 1.6$ (shape) and $b_0 = 0.4$ (rate), take a burn-in of 4000 simulations and run your code for a total of 10000 simulations when assuming your own initial values. Provide traceplots and summaries (mean, sd, 0.025, 0.5, 0.975 quantiles) of the marginal posteriors for μ, σ^2 and $1/\sigma^2$.
3. For this model, `INLA` (`https://www.r-inla.org/`) provides exact results. Compare results provided by the Gibbs sampler with those provided by `INLA`.
4. Explain step by step in your own words what the code in `04GibbsSampler.R` is doing.

Hint: If necessary, consult Ntzoufras (2009)[Chapter 2] or Robert and Casella (2010)[Chapter 7].

Exercise 3 (Metropolis-Hastings sampler for a logistic regression)

Use the file `04MHSampler.R` provided on OLAT. The code from the lecture contains a random walk Metropolis(-Hastings) sampler in `R` with one univariate normal proposal for the intercept and an independent univariate normal proposal for the slope in the logistic regression for binomial mice data (Collett, 2003, p.71). Data in Table 1 show the number of deaths from pneumonia y in n mice exposed to various doses x of an anti-pneumococcus serum. For the Bayesian analysis two independent normal priors for intercept and slope with mean $= 0$ and variance $= 10000$ are assumed.
Task: Make sure that $n.thin = 1$. Tune acceptance rates by changing the spread (tuning parameters s_alpha and s_beta) of the proposal distributions. Set the tuning parameters of the proposals to low (0.01, 1), middle (1, 100) and high (50, 5000) values to obtain different acceptance rates. Generate MCMC samples under each of the above conditions. Investigate and interpret the traceplots, auto-correlations and cross-correlations.

x	y	n
0.0028	26	28
0.0028	9	12
0.0056	21	40
0.0112	9	40
0.0225	6	40
0.0450	1	40

Table 1: Mice data from Collett (2003, p.71).

1. Explain step by step in your own words what the code in `04MHSampler.R` is doing.
2. Comment on the differences in traceplots, auto-correlations (`acf()`) and cross-correlations depending on the tuning parameters' choice. What are the reasons for observed differences?
3. Under which condition the optimal acceptance rate of about 0.2-0.4 ("rule of thumb") is attained?
4. Provide summaries (mean, sd, 0.025, 0.5, 0.975 quantiles) of the marginal posteriors for α and β for the middle choice of tuning parameters.
5. Plot the logistic curve for median posterior values of α and β together with data and interpret the result.

Hint: If necessary, consult Ntzoufras (2009)[Chapter 2] or Robert and Casella (2010)[Chapter 6].

Group contributions

Exercise 4 (Group contributions for the lecture on 28.04.2022)

Please prepare a group contribution, which your group will present (ca. 5 min) during the next lecture.

(5.1) History: WinBUGS and OpenBUGS

(5.2) History: Stan

(5.3) History: INLA

(5.4) History: bayesmeta

Make sure that the file 05contribution-Group-Name.zip (one per group) contains the pdf-file and the R code you want to present.

Optional exercises just for fun!

Exercise 5 (Two video lectures at the University of Cambridge (optional))

Listen to the following two lectures about MCMC sampling given by Professor David MacKay at the University of Cambridge:

1. Lecture 12: `videolectures.net/mackay_course_12/`
   a) 00:30:30 - 00:37:00: Motivation
   b) 00:37:00 - 00:45:05: Importance Sampling
   c) 00:45:05 - 00:53:50: Rejection Sampling
   d) 00:53:50 - 01:08:20: Metropolis
   e) 01:08:20 - 01:14:00: Gibbs
   f) 01:21:40 - 01:23:30: Comparison
2. Lecture 13: `videolectures.net/mackay_course_13/`
   a) 00:01:18 - 00:12:45: Recap
   b) 00:12:45 - 00:33:25: Random Walk
   c) 00:33:25 - 00:45:00: Dimension/Metropolis tuning
   d) 00:41:22 - 00:57:45: HMC
   e) 00:57:45 - 01:07:50: Overrelaxation
   f) 01:07:50 - 01:22:00: Slice Sampling

Alternatively, for a concise exposition of Markov Chain Monte Carlo algorithms in Bayesian inference see Ntzoufras (2009)[Chapter 2].

Exercise 6 (More challenging MCMC samplers (optional))

If you are interested in more challenging MCMC samplers have a look at Gerber and Furrer (2015).

References

Baeten, D., X. Baraliakos, J. Braun, J. Sieper, P. Emery, D. van der Heijde, I. McInnes, J. van Laar, R. Landewé, P. Wordsworth, J. Wollenhaupt, H. Kellner, J. Paramarta, J. Wei, A. Brachat, S. Bek, D. Laurent, Y. Li, Y. Wang, A. Bertolino, S. Gsteiger, A. Wright, and W. Hueber (2013). Anti-interleukin-17A monoclonal antibody secukinumab in treatment of ankylosing spondylitis: a randomised, double-blind, placebo-controlled trial. *The Lancet 382*, 1705–1713.

Collett, D. (2003). *Modelling Binary Data. Second Edition*. Chapman & Hall/CRC.

Gerber, F. and R. Furrer (2015). Pitfalls in the implementation of Bayesian hierarchical modeling of areal count data: An illustration using BYM and Leroux models. *Journal of Statistical Software 63*(Code Snippet 1), 1–32.

Ntzoufras, I. (2009). *Bayesian Modeling Using WinBUGS.* John Wiley & Sons, Inc.

Robert, C. and G. Casella (2010). *Introducing Monte Carlo Methods with R.* Springer.