TMA4300 Computer Intensive Statistical Methods

Exercise 2, Spring 2023

Note: Solutions must be handed in no later than Thursday the **23rd of March 2022, at 23:59**. All answers including derivations, computer code and graphics (all in one pdf document!) should be submitted in Blackboard.

Problem A

In this problem, we will look at a portion of the Tokyo rainfall dataset, a famous dataset with daily rainfall data from 1951–1989. We will consider the response to be whether the amount of rainfall exceeded 1mm over the given time period:

$$y_t|\tau_t \sim \text{Bin}(n_t, \pi(\tau_t)), \quad \pi(\tau_t) = \frac{\exp\{\tau_t\}}{1+\exp\{\tau_t\}} = \frac{1}{1+\exp\{-\tau_t\}},$$

for n_t being 10 for $t = 60$ (February 29th) and 39 for all other days in the year, and $\pi(\tau_t)$ being the probability of rainfall exceeding 1mm for days $t = 1, \ldots, T$ and $T = 366$. Note that τ_t is the logit probability of exceedence and can be obtained from $\pi(\tau_t)$ ($\pi(\cdot)$ is known as the 'expit' or 'inverse logit' function) via the logit function: $\tau_t = log(\pi(\tau_t)/(1-\pi(\tau_t)))$. We assume conditional independence among the $y_t|\tau_t$ for all $t = 1, \ldots, 366$.

a) Begin by downloading the Tokyo rainfall dataset from the course wiki page. Explore the dataset, plot the response as a function of t, and describe any patterns that you see.

b) Obtain the likelihood of y_t depending on parameters $\pi(\tau_t)$ for $t = 1, \ldots, 366$.

We will apply a Bayesian hierarchical model to the dataset, using a random walk of order 1 (RW(1)) to model the trend on a logit scale,

$$\tau_t \sim \tau_{t-1} + u_t,$$

for $u_t \overset{iid}{\sim} N(0, \sigma_u^2)$ so that,

$$p(\boldsymbol{\tau} \mid \sigma_u^2) = \prod_{t=2}^{T} \frac{1}{\sigma_u} \exp\left\{-\frac{1}{2\sigma_u^2}(\tau_t - \tau_{t-1})^2\right\}.$$

We will place the following inverse gamma prior on σ_u^2,

$$p(\sigma_u^2) = \frac{\beta^\alpha}{\Gamma(\alpha)}(1/\sigma_u^2)^{\alpha+1}\exp\{-\beta/\sigma_u^2\},$$

for shape α and scale β. Let $\boldsymbol{y} = (y_1 \ \dots \ y_T)^T$, $\boldsymbol{\tau} = (\tau_1 \ \dots \ \tau_T)^T$, and $\boldsymbol{\pi} = (\pi(\tau_1) \ \dots \ \pi(\tau_T))^T$.

c) Find the conditional $p(\sigma_u^2|\boldsymbol{y}, \boldsymbol{\tau})$. If the conditional is a named distribution, name it along with its associated parameters.

d) Consider the conditional prior proposal distribution, $Q(\boldsymbol{\tau}'_{\mathcal{I}}|\boldsymbol{\tau}_{-\mathcal{I}}, \sigma_u^2, \boldsymbol{y}) = p(\boldsymbol{\tau}'_{\mathcal{I}}|\boldsymbol{\tau}_{-\mathcal{I}}, \sigma_u^2)$, where $\boldsymbol{\tau}'_{\mathcal{I}}$ is the proposed values for $\boldsymbol{\tau}_{\mathcal{I}}$, $\mathcal{I} \subseteq \{1, \dots, 366\}$ is a set of time indices, and $\boldsymbol{\tau}_{-\mathcal{I}} = \boldsymbol{\tau}_{\{1,\dots,366\}\setminus\mathcal{I}}$ is $\boldsymbol{\tau}$ subset to include all indices other than those in $\mathcal{I}$. Show that the resulting acceptance probability is given by the ratio of likelihoods:

$$\alpha(\boldsymbol{\tau}'_{\mathcal{I}}|\boldsymbol{\tau}_{-\mathcal{I}}, \sigma_u^2, \boldsymbol{y}) = \min\left\{1, \ \frac{p(\boldsymbol{y}_{\mathcal{I}}|\boldsymbol{\tau}'_{\mathcal{I}})}{p(\boldsymbol{y}_{\mathcal{I}}|\boldsymbol{\tau}_{\mathcal{I}})}\right\}.$$

For the following parts of this exercise, you may use the fact that,

$$\mathrm{Prec}(\boldsymbol{\tau}|\sigma_u^2) = \frac{1}{\sigma_u^2}\mathbf{Q} = \frac{1}{\sigma_u^2}\begin{pmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & & \ddots & & \\ & & -1 & 2 & -1 \\ & & & -1 & 1 \end{pmatrix},$$

which can be derived from the above expression for $p(\boldsymbol{\tau} \mid \sigma_u^2)$. In addition, you may find the following identity for obtaining the conditional distribution of a multivariate Gaussian using precision matrices helpful. For multivariate Gaussian,

$$\boldsymbol{x} = \begin{pmatrix} \boldsymbol{x}_A \\ \boldsymbol{x}_B \end{pmatrix} \sim MVN\left(\begin{pmatrix} \boldsymbol{\mu}_A \\ \boldsymbol{\mu}_B \end{pmatrix}, \begin{pmatrix} \boldsymbol{Q}_{AA} & \boldsymbol{Q}_{AB} \\ \boldsymbol{Q}_{BA} & \boldsymbol{Q}_{BB} \end{pmatrix}^{-1}\right),$$

The conditional mean and precision for $\boldsymbol{x}_A|\boldsymbol{x}_B$ is given by:

$$\begin{aligned} \boldsymbol{\mu}_{A|B} &= \boldsymbol{\mu}_A - \boldsymbol{Q}_{AA}^{-1}\boldsymbol{Q}_{AB}(\boldsymbol{x}_B - \boldsymbol{\mu}_B) \\ \boldsymbol{Q}_{A|B} &= \boldsymbol{Q}_{AA}. \end{aligned}$$

In our application, the unit vector is a right singular vector of $\boldsymbol{Q}_{AA}^{-1}\boldsymbol{Q}_{AB}$ with associated singular value of -1. This means that under any constant mean, where $\mu_A = \mu\mathbf{1}_A$ and $\mu_B = \mu\mathbf{1}_B$, we have that:

$$\boldsymbol{\mu}_A - \boldsymbol{Q}_{AA}^{-1}\boldsymbol{Q}_{AB}(\boldsymbol{x}_B - \boldsymbol{\mu}_B) = \mu\mathbf{1}_A - \boldsymbol{Q}_{AA}^{-1}\boldsymbol{Q}_{AB}\boldsymbol{x}_B + \mu\boldsymbol{Q}_{AA}^{-1}\boldsymbol{Q}_{AB}\mathbf{1}_B) = -\boldsymbol{Q}_{AA}^{-1}\boldsymbol{Q}_{AB}\boldsymbol{x}_B.$$

We will therefore 'pretend' we have a constant mean, but henceforth ignore it in our inference, since the RW(1) model is mean invariant.

e) Implement an MCMC sampler for the posterior $p(\boldsymbol{\pi}, \sigma_u^2|\boldsymbol{y})$ using MH steps for individual τ_t parameters using the conditional prior, $p(\tau_t|\boldsymbol{\tau}_{-t}, \sigma_u)$, and Gibbs steps for σ_u^2. Assume $\alpha = 2$ and $\beta = 0.05$, which is an informative prior placing approximately 95% of the prior mass of σ_u^2 between 0.01 and 0.25. Run the MCMC algorithm for 50,000 iterations, where every element of $\boldsymbol{\tau}$ is updated per iteration, and use `proc.time()[3]` to calculate the computation time and acceptance rates. Give traceplots, histograms, and estimated autocorrelation functions for σ_u^2, $\pi(\tau_1)$, $\pi(\tau_{201})$, and $\pi(\tau_{366})$. Provide central estimates and 95% credible intervals for σ_u^2, $\pi(\tau_1)$, $\pi(\tau_{201})$, and $\pi(\tau_{366})$, and compare predictions for $\boldsymbol{\pi}$ and associated uncertainties to y_t/n_t as a function of t. Do the traceplots show the Markov chain converged? Describe your findings. (**Hint:** multiplying many probabilities together in R generally leads to poor results. Instead, it is better to work on a log scale and simplify analytically when possible, exponentiating at the end.)

f) Repeat the previous exercise, except now rather than updating individual τ_t parameters, use a conditional prior proposal involving $p(\boldsymbol{\tau}_{(a,b)}|\boldsymbol{\tau}_{-(a,b)}, \sigma_u^2)$, where $\boldsymbol{\tau}_{(a,b)} = (\tau_a \ \dots \ \tau_b)^T$ choosing intervals of length M. Explore different values of tuning parameter M, and explain your choice M. Why might incorporating a block step over the $\boldsymbol{\tau}_{(a,b)}$ parameters in this way improve efficiency of your MCMC sampler? Why might it do the opposite, depending on M? (**Hint:** precomputing $\boldsymbol{Q}_{AA}^{-1}\boldsymbol{Q}_{AB}$ as well as the Cholesky decomposition of $\boldsymbol{Q}_{AA}^{-1}$ will speed up results. You may need to do this 3 times: once for $a = 1$ and $b = M$, once for $a > 1$ and $b < 366$, and once for $b = 366$. As in *d*), every element of $\boldsymbol{\tau}$ should be updated per iteration, but when $b = 366$ you may need either a smaller block or for it to overlap with the previous block.)

Problem B

We will continue looking at the Tokyo rainfall dataset, only now using INLA rather than MCMC. INLA can be installed and loaded into R with the following code:

```
install.packages("INLA",repos=c(getOption("repos"),
INLA="https://inla.r-inla-download.org/R/stable"), dep=TRUE)
library("INLA")
```

After loading in the Tokyo rainfall dataset, we can fit the same model as in the previous problem with the following code in R:

```r
control.inla = list(strategy="simplified.laplace", int.strategy="ccd")
mod <- inla(n.rain ~ -1 + f(day, model="rw1", constr=FALSE),
            data=rain, Ntrials=n.years, control.compute=list(config = TRUE),
            family="binomial", verbose=TRUE, control.inla=control.inla)
```

Note that here we use a simplified Laplace approximation and 'ccd' integration rather than Laplace approximation and grid integration respectively. Run `?control.inla` for more information on the two options. We remove the intercept with the `-1` option, and use a RW(1) by passing the `model="rw1"` option to the `f()` function. Run `inla.doc("rw1")` for documentation provided by INLA on its built-in RW(1) model.

a) Compare the predictions and uncertainties of INLA with those of your previous Markov chains, and again use `proc.time()[3]` to calculate the computation time. Describe your findings. Note that `mod$summary.fitted.values` contains predictions and 95% CIs (what do the other 'summary' objects in `mod` contain?). Also, make sure to use the same priors as in problem 1. (**hint**: What should the prior on the intercept be? See `?control.fixed` for information on the prior for the intercept, and `?f` along with `inla.doc("rw1")` and `inla.doc(X)` where `X` is a character vector giving INLA's name for a prior for information on how to include a prior for σ_u^2. INLA places a prior on the log precision rather than the variance, so make sure to transform the prior accordingly.)

b) How robust are the results to the two `control.inla` inputs we have used? See `?control.inla`.

c) Consider the following model in INLA:

   ```r
   mod <- inla(n.rain ~ f(day, model="rw1", constr=TRUE),
               data=rain, Ntrials=n.years, control.compute=list(config = TRUE),
               family="binomial", verbose=TRUE, control.inla=control.inla)
   ```

   How is this different from the previous model mathematically, assuming the prior for σ_u^2 is set to be the same? Are its predictions significantly different, and why are they/are they not different?

Oral presentations

Date	Problem	Team
24.03	Ex2: Problem A a)-c)	Group 3
24.03	Ex2: Problem A d)-f)	Group 2
24.03	Ex2: Problem B	Group 9