TMA4300 Computer Intensive Statistical Methods
Exercise 2, Spring 2018

Note: Solutions must be handed in no later than **Tuesday March 13th, 12:00**. All answers including derivations, computer code and graphics (all in one pdf document!) should be submitted in Blackboard as specified in the course home page.

Getting started: The aim of this exercise is to get experience with MCMC algorithms by implementing and running such algorithms for various target distributions, and to use the MCMC output to estimate properties of the target distributions. Whenever you need to sample from a standard univariate distribution you may use the build-in random number functions in R or you can use the corresponding functions you coded in Exercise 1. However, remember that there are two common parameterisations for the gamma distribution, so if you use the build-in function *rgamma* you must first check what parameterisation is used by this function.

Important: For each function of code chunk you write in this exercise you should try to check that it is working properly and describe the procedure and include plots and numerical values you use for this in your solution text. You typically do not have available analytical properties for the target distribution as in Exercise 1, but you should still carefully consider whether the simulated values are reasonable.

Whenever you are using MCMC output to estimate properties of the target distribution you need to discuss how you decided the length of the burn-in period and include the relevant plots in your solution.

Problem A: The coal-mining disaster data

In this problem you should analyse a famous data set of time intervals between successive coal-mining disasters in the UK involving ten or more men killed (Jarrett, 1979). The data is for the period March 15th 1851 to March 22nd 1962. In R this data set becomes available by running the command 'library(boot)'. The data set is then in the variable 'coal'. You should note that the first and last records in this variable are the start and end dates, respectively. In the period there were thereby 189 coal-mining disasters.

1. Get a first impression of the data set by making a plot with year along the x-axis and cumulative number of disasters along the y-axis. Discuss what you can learn from this figure.

We adopt a hierarchical Bayesian model to analyse the data set. We assume the coal-mining disasters to follow an inhomogeneous Poisson process with intensity function $\lambda(t)$ (number of events per year). We assume $\lambda(t)$ to be piecewise constant with n breakpoints. We let t_0 and t_{n+1} denote the start and end times for the data set and let $t_k; k = 1, \ldots, n$ denote the break points of the intensity function. Thus,

$$\lambda(t) = \begin{cases} \lambda_{k-1} & \text{for } t \in [t_{k-1}, t_k),\ k = 1, \ldots, n, \\ \lambda_n & \text{for } t \in [t_n, t_{n+1}]. \end{cases} \tag{1}$$

The parameters of this model is thereby $t_1, \ldots, t_n$ and $\lambda_0, \ldots, \lambda_n$, where $t_0 < t_1 < \cdots < t_n < t_{n+1}$. By subdividing the observation period into short intervals, and taking the limit when the length of these intervals go to zero, one can derive the likelihood function for the observed data as

$$f(x|t_1, \ldots, t_n, \lambda_1, \ldots, \lambda_n) = \exp\left(-\int_{t_0}^{t_{n+1}} \lambda(t)\mathrm{d}t\right) \prod_{k=0}^{n} \lambda_k^{y_k} = \exp\left(-\sum_{k=0}^{n} \lambda_k (t_{k+1} - t_k)\right) \prod_{k=0}^{n} \lambda_k^{y_k}, \tag{2}$$

where x is the observed data and y_k is the number of observed disasters in the period t_k to t_{k+1}. Assume $t_1, \ldots, t_n$ to be apriori uniformly distributed on the allowed values, and $\lambda_0, \ldots, \lambda_n$ to be apriori independent of $t_1, \ldots, t_n$ and apriori independent of each other. Apriori we assume all $\lambda_0, \ldots, \lambda_n$ to be distributed from the same gamma distribution with shape parameter $\alpha = 2$ and scale parameter β, i.e.

$$f(\lambda_i|\beta) = \frac{1}{\beta^2}\lambda_i e^{-\frac{\lambda_i}{\beta}} \quad \text{for } \lambda_i \geq 0. \tag{3}$$

Finally, for β we use the improper prior

$$f(\beta) \propto \frac{\exp\{-\frac{1}{\beta}\}}{\beta} \quad \text{for } \beta > 0. \tag{4}$$

In the following we only assume $n = 1$, so the model parameters are $\theta = (t_1, \lambda_0, \lambda_1, \beta)$.

2. Find an expression for the posterior distribution (up to a normalising constant) for θ given x, $f(\theta|x)$.

3. Find the full conditionals for each of the elements in θ. Identify all except one of these as belonging to named distributions.

4. Define and implement a single site MCMC algorithm for $f(\theta|x)$. Specify in particular what proposal distributions you are using and give formulas for the corresponding Metropolis–Hastings acceptance probabilities.

5. Run the algorithm. Evaluate the burn-in and mixing properties of your algorithm. *Hint: Use the plot you generated in 1 to evaluate whether the simulated values are reasonable.*

6. If your algorithm has a tuning parameter, explore how its value influence the length of the burn-in and the mixing properties of the simulated Markov chain, and demonstrate that its value does not influence the limiting distribution.

7. Define and implement a block Metropolis–Hastings algorithm for $f(\theta|x)$ using the following two (block) proposals.

   (a) A block proposal for $(t_1, \lambda_0, \lambda_1)$ keeping β unchanged. Generate the potential new values $(\widetilde{t}_1, \widetilde{\lambda}_0, \widetilde{\lambda}_1)$ by first generating $\widetilde{t}_1$ from a normal distribution centered at the current value of t_1 and thereafter generate $(\widetilde{\lambda}_0, \widetilde{\lambda}_1)$ from their joint full conditionals inserted the potential new value $\widetilde{t}_1$, i.e. $f(\lambda_1, \lambda_2|x, \widetilde{t}_1, \beta)$.

   (b) A block proposal for $(\beta, \lambda_0, \lambda_1)$ keeping t_1 unchanged. Generate the potential new values $(\widetilde{\beta}, \widetilde{\lambda}_0, \widetilde{\lambda}_2)$ by first generating $\widetilde{\beta}$ from a normal distribution centered at the current value β and thereafter generate $(\widetilde{\lambda}_0, \widetilde{\lambda}_1)$ from their resulting joint full conditional inserted $\widetilde{\beta}$, i.e. $f(\lambda_0, \lambda_1|x, t_1, \widetilde{\beta})$.

8. Run the new algorithm for different values of the tuning parameters and evaluate the burn-in and mixing properties of the algorithm. Compare the burn-in and mixing properties of the single site and block Metropolis–Hastings algorithms. Use the simulation results to estimate the marginal posterior distributions $f(t_1|x), f(\lambda_0|x), f(\lambda_1|x)$ and $f(\beta|x)$. Use also the simulated values to estimate $\mathrm{E}[t_1|x]$, $\mathrm{E}[\lambda_0|x]$, $\mathrm{E}[\lambda_1|x]$ and $\mathrm{Cov}[\lambda_0, \lambda_1|x]$.

Problem B: **Bayesian image reconstruction**

In this problem we will study how the Ising model can be used as a prior distribution in an image reconstruction setting. Let $y = (y_{ij}, i = 1, \ldots, 89, j = 1, \ldots, 85)$ be the observed "image.txt" you can download from the course web page. We assume this to be a noisy version of an unobserved binary image $x = (x_{ij}, i = 1, \ldots, 89, j = 1, \ldots, 85)$ with $x_{ij} \in \{0, 1\}$. Our goal in this problem is to use the observed y to estimate x. We assume the elements in y to be conditionally independent given x, and

$$y_{ij}|x \sim N(\mu_{x_{ij}}, \sigma^2_{x_{ij}}), \tag{5}$$

where μ_0, μ_1 are the mean values for y_{ij} when x_{ij} is zero and one, respectively, and σ_0^2 and σ_1^2 are corresponding variances. Apriori we assume x to be independent of $\varphi = (\mu_0, \mu_1, \sigma_0, \sigma_2)$. As prior for x we assume an Ising model with interaction parameter β, i.e.

$$f(x) \propto \exp\left\{ \beta \sum_{(i,j)\sim(k,l)} I(x_{ij} = x_{kl}) \right\}, \tag{6}$$

where the sum is over all pairs of neighbour nodes in the 89×85 lattice and the value of β is assumed to be known. To define a prior for φ we follow a procedure used in Austad and Tjelmeland (2017). We first define a reparameterisation to new parameters (m_0, θ, s, τ) by the relations

$$\sigma_0 = s \cdot \tau, \quad \sigma_1 = \frac{s}{\tau}, \quad \mu_0 = m_0, \quad \text{and} \quad \mu_1 = m_0 + s\theta. \tag{7}$$

The s defines a scale, θ defines the difference between the two mean values in this scale, and τ defines σ_0 and σ_1 using the same scale. We then define a prior for $\varphi = (\mu_0, \mu_1, \sigma_0, \sigma_1)$ implicitly by assigning a prior for (m_0, θ, s, τ). Apriori we assume the four parameters m_0, θ, s and τ to be independent. For m_0 and s we assume improper uniform distributions on $(-\infty, \infty)$ and $(0, \infty)$, respectively. For θ we assume a gamma prior with mean and variance both equal to four. Finally, as prior for τ we use

$$f(\tau) = \begin{cases} \frac{1}{2\tau^2} e^{-\left(\frac{1}{\tau} - 1\right)} & \text{for } \tau \in (0, 1], \\ \frac{1}{2} e^{-(\tau - 1)} & \text{for } \tau > 1. \end{cases} \tag{8}$$

1. From the prior for τ defined in (8), use the transformation formula to find the corresponding prior for $t = 1/\tau$. Show that the priors for τ and $t = 1/\tau$ are identical and use this to argue that therefore also the (marginal) priors for σ_0 and σ_1 are identical.

2. Show that the resulting (improper) prior for $\varphi = (\mu_0, \mu_1, \sigma_0, \sigma_1)$ becomes (up to proportionality)

$$f(\varphi) \propto \begin{cases} \frac{(\mu_1 - \mu_0)^3}{\sigma_0^3 \sigma_1^2} \exp\left\{ -\left[\frac{\mu_1 - \mu_0}{\sqrt{\sigma_0 \sigma_1}} + \sqrt{\frac{\sigma_1}{\sigma_0}} \right] \right\} & \text{for } \sigma_0 \leq \sigma_1 \text{ and } \mu_0 < \mu_1, \\ \frac{(\mu_1 - \mu_0)^3}{\sigma_0^2 \sigma_1^3} \exp\left\{ -\left[\frac{\mu_1 - \mu_0}{\sqrt{\sigma_0 \sigma_1}} + \sqrt{\frac{\sigma_0}{\sigma_1}} \right] \right\} & \text{for } \sigma_0 > \sigma_1 \text{ and } \mu_0 < \mu_1. \end{cases} \tag{9}$$

3. Find (up to proportionality) a formula for the posterior distribution $f(x, \varphi|y)$

4. Define and implement a Metropolis–Hastings algorithm for simulating from $f(x, \varphi|y)$, using the following updates.

   (a) For a randomly chosen node (i, j), propose a potential new value for x_{ij} by deterministically setting $\widetilde{x}_{ij} = 1 - x_{ij}$. Accept the proposal with the usual Metropolis–Hastings acceptance probability.

   (b) Propose a potential new value for μ_0 from the resulting full conditional when assuming an improper flat prior. Accept the proposal with the usual Metropolis–Hastings acceptance probability.

   (c) Propose a potential new value for μ_1 from the resulting full conditional when assuming an improper flat prior. Accept the proposal with the usual Metropolis–Hastings acceptance probability.

(d) Propose a potential new value for σ_0 by simulating σ_0^2 from the resulting full conditional for σ_0^2 when assuming an inverse gamma prior with large variance for σ_0^2. Accept the proposal with the usual Metropolis–Hastings acceptance probability.

(e) Propose a potential new value for σ_1 by simulating σ_1^2 from the resulting full conditional for σ_1^2 when assuming an inverse gamma prior with large variance for σ_1^2. Accept the proposal with the usual Metropolis–Hastings acceptance probability.

For each of the updates, give formulas for the proposal distribution and the acceptance probability. Define for example one iteration of the Metropolis–Hastings algorithm to consists of $89 \cdot 85$ updates of type (a) and one update for each of the other four update types.

5. Run the Markov chain defined above for $\beta = 0$, for $\beta = 0.6$ and for $\beta = 1.0$. Evaluate the burn-in and mixing properties of your algorithm. What is the acceptance rates for each of the five update types? *Hint: To evaluate whether the simulated $\mu_0, \mu_1, \sigma_0, \sigma_1$ values are reasonable compare with the histogram of all the observed y_{ij} values.*

6. Use the simulation output to estimate the following properties when $\beta = 0$, when $\beta = 0.6$ and when $\beta = 1.0$.

   (a) For each (i, j) in the lattice: $\mathrm{E}[x_{ij}|y]$. Visualise the estimates as an image. Use these numbers to estimate also the posterior marginal most probable value for each x_{ij},

   $$\arg\max_{x_{ij}} f(x_{ij}|y). \tag{10}$$

   Again visualise your estimates as an image. For what fraction of the nodes is the posterior most probable value for x_{ij} equal to one?

   (b) The expected fraction of nodes with $x_{ij} = 1$, i.e.

   $$\gamma = \mathrm{E}\left[\frac{1}{89 \cdot 85} \sum_{i=1}^{89} \sum_{j=1}^{85} x_{ij} \,\middle|\, y \right], \tag{11}$$

   where the sum is over all nodes in the lattice. Compare with your results in (a) and discuss.

7. Use your simulation results for $\beta = 1.0$ to estimate also the following properties.

   (a) The marginal distributions for each of μ_0, μ_1, σ_0 and σ_1.

   (b) For γ defined in (11), estimate also the variance of your estimate and use this to find a 90% confidence interval for γ.

   (c) A 90% credible interval for the fraction of nodes with $x_{ij} = 1$. Compare with your confidence interval in (b) and discuss.

References

Austad, H. and Tjelmeland, H. (2017). Approximate computations for binary Markov random fields and their use in Bayesian models. Statistics and Computing, **27**, 1271–1292.

Jarett, R.G. (1979). A note on the intervals between coal-mining disasters. Biometrika, **66**, 191–193.

Oral presentations

Date	Problem	Team
09.03.2018	2: Problems A1 to A4	Sigrid Leithe and Norunn Ahdell Wankel
	Ask at least one question:	Kristoffer Skuland and Sindre Nybakk Uthus
09.03.2018	2: Problems A5 to A7	Kwaku Peprah Adjel and Magnus Liland
	Ask at least one question:	Marcus Aleksander Engebretsen and Gina Magnussen
09.03.2018	2: Problems B1 to B3	Jens Andreas Teigland Holck and Sindre Aalvik Stranden
	Ask at least one question:	Martin André Brevik Blindheimsvik and Morten Grønbech
09.03.2018	2: Problems B4 and B5	Elisabeth Hetlelid and Dag Johnsrud Kristiansen
	Ask at least one question:	Mira Lilleholt Vik
09.03.2017	2: Problems B6 and B7	Yi Liu and Fredrik Nevjen
	Ask at least one question:	Thomas Benjamin Frogner and Rasmus Andreas Wichstrøm Münter