Project 2

Marcus Hjørund & Gard Gravdal

2024-03-15

Contents

1 Problem 1

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

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

for n_t being 10 for $t = 60$ (Feburary 29th) and 39 for all other days. $\pi(x_t)$ is the probability of rainfall exceeding 1mm of days $t = 1, \ldots, T$ and $T = 366$. Note that x_t is the logit probability of exceedence and can be obtained from $\pi(x_t)$ via $x_t = \log(\pi(x_t)/(1-\pi(x_t)))$. We assume conditional independence among the $y_t|x_t$ for all $t = 1, \ldots, 366$.

1.1 a)

As we can see in Figure 1 the rain in Tokyo seems to show seasonal trend. There are few occurences of rain more than 1mm in the period December-January, the peak of rainfall is in June-July, with a following drier period in August-September.

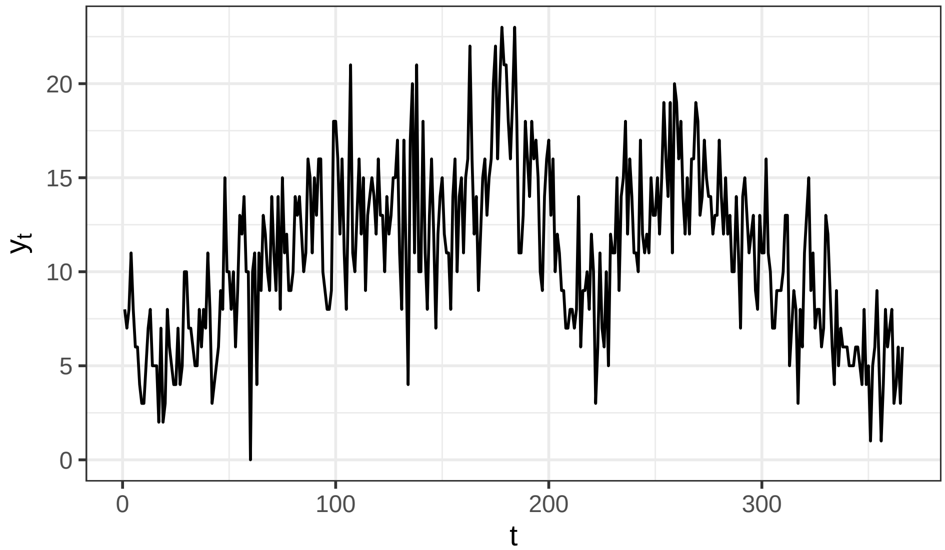

Figure 1: Tokyo rain data

1.2 b)

Next we want to obtain an expression for the likelihood of y_t depending on the parameters $\pi(x_t)$ for $t = 1, \ldots, 366$. This gives the likeihood for the data $\boldsymbol{y}$ given the probabilities $\pi(\boldsymbol{x})$, where we will later update $\boldsymbol{x}$ using our a MCMC algorithm. Since we assume conditional intependence for $y_t|x_t$ we can write the likehood, using the expression for $\pi(x_t)$ given in Equation (1):

$$
\begin{aligned}
L(\pi(\boldsymbol{x})|\boldsymbol{y}) &= \prod_{t=1}^{T} \binom{n_t}{y_t} \left(\frac{\exp(x_t)}{1+\exp(x_t)} \right)^{y_t} \left(1 - \frac{\exp(x_t)}{1+\exp(x_t)} \right)^{n_t - y_t} \\
&= \prod_{t=1}^{T} \binom{n_t}{y_t} \left(\frac{\exp(x_t)}{1+\exp(x_t)} \right)^{y_t} \left(\frac{1}{1+\exp(x_t)} \right)^{n_t - y_t} \\
&= \prod_{t=1}^{T} \binom{n_t}{y_t} \frac{\exp(x_t y_t)}{(1+\exp(x_t))^{n_t}} \qquad (2)
\end{aligned}
$$

1.3 c)

Now we want to apply a Bayesian hierarchical model to the dataset, where we use a random walk of order 1 (RW(1)) to model the trend on a logit scale,

$$x_t = x_{t-1} + u_t,$$

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

$$p(\boldsymbol{x}|\sigma_u^2) \propto \prod_{t=2}^{T} \frac{1}{\sigma_u} \exp\left\{ -\frac{1}{2\sigma_u^2}(x_t - x_{t-1})^2 \right\}. \qquad (3)$$

Then we place a inverse gamma prior on σ_u^2 such that,

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

for shape and scale α and β. We let $\boldsymbol{y} = (y_1, y_2, \ldots, y_T)^T$, $\boldsymbol{x} = (x_1, x_2, \ldots, x_T)^T$ and $\boldsymbol{\pi} = (\pi(x_1), \pi(x_2), \ldots, \pi(x_T))^T$.

Now we will derive the posteriori distribution for $\sigma_u^2|\boldsymbol{x}$, which is used in Gibbs sampling in the hybrid sampler in Section 1.6. If we draw a directed acyclical graph for the hierarchical Bayesian model, we see that $\boldsymbol{y}$ is independent of σ_u^2, as all information from $\boldsymbol{y}$ is stored in $\boldsymbol{x}$. From Bayes theorem we know that

$$p(\sigma_u^2|\boldsymbol{y}, \boldsymbol{x}) \propto p(\sigma_u^2)p(\boldsymbol{x}|\sigma_u^2).$$

We then have that

$$\begin{aligned}
p(\sigma_u^2|\boldsymbol{x}, \boldsymbol{y}) &\propto \frac{\beta^\alpha}{\Gamma(\alpha)}(1/\sigma_u^2)^{\alpha+1}\exp(-\beta/\sigma_u^2)\prod_{t=2}^{T}\frac{1}{\sigma_u}\exp\left\{-\frac{1}{2\sigma_u^2}(x_t - x_{t-1})^2\right\} \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)}(1/\sigma_u^2)^{\alpha+1}\exp(-\beta/\sigma_u^2)\cdot(\sigma_u^2)^{-\frac{T-1}{2}}\exp\left\{-\frac{1}{2\sigma_u^2}\sum_{t=2}^{T}(x_t - x_{t-1})^2\right\} \\
&\propto (\sigma_u^2)^{-(\alpha+\frac{T-1}{2}+1)}\exp\left\{-\frac{1}{\sigma_u^2}\left(\beta + \frac{1}{2}\sum_{t=2}^{T}(x_t - x_{t-1})^2\right)\right\} \qquad (4)
\end{aligned}$$

which follows an inverse gamma distribution with parameters

$$\alpha^* = \alpha + \frac{T-1}{2} \text{ and } \beta^* = \beta + \frac{1}{2}\sum_{t=2}^{T}(x_t - x_{t-1})^2.$$

1.4 d)

In the Metropolis-Hastings algorithm, which we will implement, we need to find an expression for the acceptance probability α, which in general is given as

$$\alpha(y|x) = \min\left\{1, \frac{\pi(y)Q(x|y)}{\pi(x)Q(y|x)}\right\}$$

where $Q(x|y)$ is a proposed conditional probability mass function (pmf), $\pi(x)$ is the probability of being in state x, and the transition probability is given as

$$p(y|x) = \begin{cases} Q(y|x)\alpha(y|x) & \text{for } y \neq x \\ 1 - \sum_{y\neq x} Q(y|x)\alpha(y|x) & \text{for } y = x \end{cases}$$

which is the probability of transitioning to state y given that the current state is x. The expression for α can be derived from the detailed balance equation, $\pi(x)p(y|x) = \pi(y)p(x|y)$ which is true for a time reversible Markov chain.

We will now consider the conditional prior proposal distribution, $Q(\boldsymbol{x}'_{\mathcal{I}}|\boldsymbol{x}_{-\mathcal{I}}, \sigma_u^2, \boldsymbol{y}) = p(\boldsymbol{x}'_{\mathcal{I}}|\boldsymbol{x}_{-\mathcal{I}}, \sigma_u^2)$, where $\boldsymbol{x}'_{\mathcal{I}}$ is the proposed values for $\boldsymbol{x}_{\mathcal{I}}$, $\mathcal{I} \subseteq \{1, \ldots, 366\}$ is a set of time indices, and $\boldsymbol{x}_{-\mathcal{I}} = \boldsymbol{x}_{\{1,\ldots,366\}\setminus\mathcal{I}}$ is a subset of $\boldsymbol{x}$ that includes all other indices than those in $\mathcal{I}$. Our expression for the acceptance probability is then

$$\begin{aligned}
\alpha(\boldsymbol{x}'_{\mathcal{I}}|\boldsymbol{x}_{-\mathcal{I}}, \sigma_u^2, \boldsymbol{y}) &= \min\left\{1, \frac{\pi(\boldsymbol{x}'_{\mathcal{I}}|\boldsymbol{x}_{-\mathcal{I}}, \sigma_u^2, \boldsymbol{y})Q(\boldsymbol{x}_{\mathcal{I}}|\boldsymbol{x}'_{\mathcal{I}}, \boldsymbol{x}_{-\mathcal{I}}, \sigma_u^2, \boldsymbol{y})}{\pi(\boldsymbol{x}_{\mathcal{I}}|\boldsymbol{x}_{-\mathcal{I}}, \sigma_u^2, \boldsymbol{y})Q(\boldsymbol{x}'_{\mathcal{I}}|\boldsymbol{x}_{\mathcal{I}}, \boldsymbol{x}_{-\mathcal{I}}, \sigma_u^2, \boldsymbol{y})}\right\} \\
&= \min\left\{1, \frac{p(\boldsymbol{x}'_{\mathcal{I}}|\boldsymbol{x}_{-\mathcal{I}}, \sigma_u^2, \boldsymbol{y})p(\boldsymbol{x}_{\mathcal{I}}|\boldsymbol{x}_{-\mathcal{I}}, \sigma_u^2)}{p(\boldsymbol{x}_{\mathcal{I}}|\boldsymbol{x}_{-\mathcal{I}}, \sigma_u^2, \boldsymbol{y})p(\boldsymbol{x}'_{\mathcal{I}}|\boldsymbol{x}_{-\mathcal{I}}, \sigma_u^2)}\right\}.
\end{aligned}$$

We can simplify this expression by looking at $p(\boldsymbol{x}_{\mathcal{I}}|\boldsymbol{x}_{-\mathcal{I}}, \sigma_u^2, \boldsymbol{y})$ and $p(\boldsymbol{x}'_{\mathcal{I}}|\boldsymbol{x}_{-\mathcal{I}}, \sigma_u^2, \boldsymbol{y})$. From probability theory we know that $P(A|B) = \frac{P(A\cap B)}{P(B)}$, such that we can divide the expression of $p(\boldsymbol{x}'_{\mathcal{I}}|\boldsymbol{x}_{-\mathcal{I}}, \sigma_u^2, \boldsymbol{y})$ as follows:

$$\begin{aligned} p(\boldsymbol{x}'_{\mathcal{I}}|\boldsymbol{x}_{-\mathcal{I}}, \sigma_u^2, \boldsymbol{y}) &= \frac{p(\boldsymbol{x}'_{\mathcal{I}}, \boldsymbol{y}|\boldsymbol{x}_{-\mathcal{I}}, \sigma_u^2)}{p(\boldsymbol{y}|\boldsymbol{x}_{-\mathcal{I}}, \sigma_u^2)} = \frac{p(\boldsymbol{y}|\boldsymbol{x}'_{\mathcal{I}}, \boldsymbol{x}_{-\mathcal{I}}, \sigma_u^2)\cdot p(\boldsymbol{x}'_{\mathcal{I}}|\boldsymbol{x}_{-\mathcal{I}}, \sigma_u^2)}{p(\boldsymbol{y}|\boldsymbol{x}_{-\mathcal{I}}, \sigma_u^2)} \\ &= \frac{p(\boldsymbol{y}_{\mathcal{I}}|\boldsymbol{x}'_{\mathcal{I}})\cdot p(\boldsymbol{y}_{-\mathcal{I}}|\boldsymbol{x}_{-\mathcal{I}})\cdot p(\boldsymbol{x}'_{\mathcal{I}}|\boldsymbol{x}_{-\mathcal{I}}, \sigma_u^2)}{p(\boldsymbol{y}|\boldsymbol{x}_{-\mathcal{I}})} \end{aligned}$$

where we have used that $\boldsymbol{y}$ is independent of σ_u^2 and that $\boldsymbol{x}_{\mathcal{I}}$ and $\boldsymbol{x}_{-\mathcal{I}}$ are disjoint sets, and $y_t|x_t$ are conditionally independent. We perform a similar calculation for $p(\boldsymbol{x}_{\mathcal{I}}|\boldsymbol{x}_{-\mathcal{I}}, \sigma_u^2, \boldsymbol{y})$ and obtain

$$p(\boldsymbol{x}_{\mathcal{I}}|\boldsymbol{x}_{-\mathcal{I}}, \sigma_u^2, \boldsymbol{y}) = \frac{p(\boldsymbol{y}_{\mathcal{I}}|\boldsymbol{x}_{\mathcal{I}})\cdot p(\boldsymbol{y}_{-\mathcal{I}}|\boldsymbol{x}_{-\mathcal{I}})\cdot p(\boldsymbol{x}_{\mathcal{I}}|\boldsymbol{x}_{-\mathcal{I}}, \sigma_u^2)}{p(\boldsymbol{y}|\boldsymbol{x}_{-\mathcal{I}})}.$$

We can then simplify $\alpha(\boldsymbol{x}'_{\mathcal{I}}|\boldsymbol{x}_{-\mathcal{I}}, \sigma_u^2, \boldsymbol{y})$, which becomes

$$\alpha(\boldsymbol{x}'_{\mathcal{I}}|\boldsymbol{x}_{-\mathcal{I}}, \sigma_u^2, \boldsymbol{y}) = \min\left\{1, \frac{p(\boldsymbol{x}'_{\mathcal{I}}|\boldsymbol{x}_{-\mathcal{I}}, \sigma_u^2, \boldsymbol{y})p(\boldsymbol{x}_{\mathcal{I}}|\boldsymbol{x}_{-\mathcal{I}}, \sigma_u^2)}{p(\boldsymbol{x}_{\mathcal{I}}|\boldsymbol{x}_{-\mathcal{I}}, \sigma_u^2, \boldsymbol{y})p(\boldsymbol{x}'_{\mathcal{I}}|\boldsymbol{x}_{-\mathcal{I}}, \sigma_u^2)}\right\} = \min\left\{1, \frac{p(\boldsymbol{y}_{\mathcal{I}}|\boldsymbol{x}'_{\mathcal{I}})}{p(\boldsymbol{y}_{\mathcal{I}}|\boldsymbol{x}_{\mathcal{I}})}\right\}. \tag{5}$$

We can then see that $\alpha(\boldsymbol{x}'_{\mathcal{I}}|\boldsymbol{x}_{-\mathcal{I}}, \sigma_u^2, \boldsymbol{y})$ becomes a ratio of likelihoods for the proposed values $\boldsymbol{x}'_{\mathcal{I}}$ against the existing values $\boldsymbol{x}_{\mathcal{I}}$.

1.5 e)

We note that the density specified by (3) is improper, for example for $T = 2$, density takes the shape of an infinite 'Gaussian ridge' centered around the line given by $x_1 = x_2$. Equation (3) can be rewritten as

$$p(\boldsymbol{x}|\sigma_u^2) \propto \exp\left\{-\frac{1}{2}\boldsymbol{x}^T\mathbf{Q}\boldsymbol{x}\right\} \tag{6}$$

which resembles a multivariate normal density but where the improperiety of the density translates into the precision matrix

$$\mathbf{Q} = \frac{1}{\sigma_u^2}\begin{pmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & & \ddots & & \\ & & -1 & 2 & -1 \\ & & & -1 & 1 \end{pmatrix},$$

having one zero eigenvalue. We partition the components of $\boldsymbol{x}$ into two subvectors writing

$$\boldsymbol{x} = \begin{pmatrix} \boldsymbol{x}_A \\ \boldsymbol{x}_B \end{pmatrix},$$

and partitioning the precision matrix in the same way as

$$\mathbf{Q} = \begin{pmatrix} \mathbf{Q}_{AA} & \mathbf{Q}_{AB} \\ \mathbf{Q}_{BA} & \mathbf{Q}_{BB} \end{pmatrix}.$$

We then want to derive the conditional distribution of $\boldsymbol{x}_A$ conditional on $\boldsymbol{x}_B$. We know from before that the conditional density is always proportional to the joint (improper) density. We assume without loss of generality that $\boldsymbol{\mu} = \mathbf{0}$ and find that

$$
\begin{aligned}
f_{\boldsymbol{x}_A|\boldsymbol{x}_B}(\boldsymbol{x}_A) &\propto \exp\left\{-\frac{1}{2}(\boldsymbol{x}_A^T, \boldsymbol{x}_B^T)\begin{bmatrix}\mathbf{Q}_{AA} & \mathbf{Q}_{AB} \\ \mathbf{Q}_{BA} & \mathbf{Q}_{BB}\end{bmatrix}\begin{pmatrix}\boldsymbol{x}_A \\ \boldsymbol{x}_B\end{pmatrix}\right\} \\
&= \exp\left\{-\frac{1}{2}(\boldsymbol{x}_A^T\mathbf{Q}_{AA}\boldsymbol{x}_A + \boldsymbol{x}_B^T\mathbf{Q}_{BA}\boldsymbol{x}_A + \boldsymbol{x}_A^T\mathbf{Q}_{AB}\boldsymbol{x}_B + \boldsymbol{x}_B^T\mathbf{Q}_{BB}\boldsymbol{x}_B)\right\} \\
&\propto \exp\left\{-\frac{1}{2}(\boldsymbol{x}_A - \boldsymbol{\mu}_{A|B})^T\mathbf{Q}_{AA}(\boldsymbol{x}_B - \boldsymbol{\mu}_{A|B})\right\} \\
&\propto \exp\left\{-\frac{1}{2}(\boldsymbol{x}_A^T\mathbf{Q}_{AA}\boldsymbol{x}_A - \boldsymbol{\mu}_{A|B}^T\mathbf{Q}_{AA}\boldsymbol{x}_A - \boldsymbol{x}_A^T\mathbf{Q}_{AA}\boldsymbol{\mu}_{A|B})\right\}
\end{aligned}
$$

where we only consider terms involving $\boldsymbol{x}_A$, which is random in this case, and we use the fact that we can write the terms involving $\boldsymbol{x}_A$ in the second line in quadratic form, completing the square. We then equate the coefficients in the last line to the terms in the second line to find our expressions for $\boldsymbol{\mu}_{A|B}$ and $\mathbf{Q}_{A|B}$. We have

$$
-\mathbf{Q}_{AA}\boldsymbol{\mu}_{A|B} = \mathbf{Q}_{AB}\boldsymbol{x}_B \implies \boldsymbol{\mu}_{A|B} = -\mathbf{Q}_{AA}^{-1}\mathbf{Q}_{AB}\boldsymbol{x}_B \tag{7}
$$

$$
\mathbf{Q}_{A|B} = \mathbf{Q}_{AA}. \tag{8}
$$

These results is used to simulate from $\boldsymbol{x}_{\mathcal{I}}'|\boldsymbol{x}_{-\mathcal{I}}$ later on when we generate realizations from the random walk.

1.6 f)

In this section we want to implement an MCMC sampler from the posterior distribution $p(\boldsymbol{\pi}, \sigma_u^2|\boldsymbol{y})$ using Metropolis-Hastings steps for individual x_t parameters using the conditional prior, $p(x_t|\boldsymbol{x}_{-t}\sigma_u^2)$, and Gibbs steps for σ_u^2.

We will begin by deriving the conditional prior distribution using the results from Equations (7) and (8) as the conditional mean and precision (inverse variance) of $x_{\mathcal{I}}'|\boldsymbol{x}_{-\mathcal{I}}$. We need to derive the conditional prior distribution for three cases, $t = 1$, $2 \leq t \leq 365$ and $t = 366$. For the case $t = 1$, $\mathcal{I} = \{1\}$, we have for the conditional mean that $-\mathbf{Q}_{\mathcal{I}\mathcal{I}}^{-1} = -\sigma_u^2$, $\mathbf{Q}_{\mathcal{I}-\mathcal{I}} = \frac{1}{\sigma_u^2}(-1, 0, \ldots, 0)$ with dimension $1 \times (T-1)$. Thus the conditional expectation is given as

$$
x_{\mathcal{I}}|\boldsymbol{x}_{-\mathcal{I}} = -\mathbf{Q}_{\mathcal{I}\mathcal{I}}^{-1}\mathbf{Q}_{\mathcal{I}-\mathcal{I}}\boldsymbol{x}_{-\mathcal{I}} = -\sigma_u^2 \cdot \frac{1}{\sigma_u^2}(-1, 0, \ldots, 0)(x_2, x_3, \ldots, x_T)^T = x_2. \tag{9}
$$

Similarly we have for $t = 366$, $\mathcal{I} = \{366\}$, we have

$$
x_{\mathcal{I}}|\boldsymbol{x}_{-\mathcal{I}} = -\sigma_u^2 \cdot \frac{1}{\sigma_u^2}(0, \ldots, 0, -1)(x_1, \ldots, x_{364}, x_{365})^T = x_{365}. \tag{10}
$$

Then finally for the interior points, $2 \leq t \leq 365$, $\mathcal{I} = \{2\}, \{3\}, \ldots, \{365\}$, we have $-\mathbf{Q}_{\mathcal{I}\mathcal{I}}^{-1} = -\frac{\sigma_u^2}{2}$ and $\mathbf{Q}_{\mathcal{I}-\mathcal{I}} = \frac{1}{\sigma_u^2}(0, \ldots, 0, -1, -1, 0, \ldots, 0)$ with dimension $1 \times (T-1)$. We then calculate the conditional expectation as

$$
x_{\mathcal{I}}|\boldsymbol{x}_{-\mathcal{I}} = -\frac{\sigma_u^2}{2} \cdot \frac{1}{\sigma_u^2}(0, \ldots, 0, -1, -1, 0, \ldots, 0)(x_1, \ldots, x_{t-1}, x_{t+1}, \ldots, x_{366})^T = \frac{1}{2}(x_{t-1} + x_{t+1}). \tag{11}
$$

We perform a similar calculation to calculate the conditional variance. We have that the conditional variance $\boldsymbol{\Sigma}_{\mathcal{I}|-\mathcal{I}} = \mathbf{Q}_{\mathcal{I}|-\mathcal{I}}^{-1}$. For $t = 1$ and $t = 366$ we have that $\mathbf{Q}_{\mathcal{I}\mathcal{I}}^{-1} = \sigma_u^2$ and for $2 \leq t \leq 365$ we have $\mathbf{Q}_{\mathcal{I}\mathcal{I}}^{-1} = \frac{\sigma_u^2}{2}$. Thus we can summarize the conditional distribution of $x_{\mathcal{I}}|\boldsymbol{x}_{-\mathcal{I}}$ as

$$
x_{\mathcal{I}}|\boldsymbol{x}_{-\mathcal{I}} \sim \begin{cases} \mathcal{N}(x_2, \sigma_u^2) & \text{for } t = 1 \\ \mathcal{N}(\frac{1}{2}(x_{t-1} + x_{t+1}), \frac{\sigma_u^2}{2}) & \text{for } t = 2, \ldots, 365 \\ \mathcal{N}(x_{365}, \sigma_u^2) & \text{for } t = 366. \end{cases} \tag{12}
$$