Solution Sketch for TMA4300

Spring 2020

Problem 1

1. One possibility could be to use an exponential with parameter 2 as proposal distribution $g(x)$. One would then have that
$$\frac{f(x)}{g(x)} = \frac{Ce^{-2x}|\sin(x)|}{\exp\{-2x\}} = C|\sin(x)| \leq C$$
Since $|\sin(x)|$ take values in (0,1) for $x \in \mathcal{R}^+$. The rejection sampling algorithm is then:
    1. Simulate $x \sim \exp(1)$
    2. Simulate $u \sim \text{Unif}(0,1)$
    3. Compute $\alpha = \frac{1}{C}C\exp(-x)|\sin(x)| = \exp(-x)|\sin(x)|$
    4. If $u < \alpha$ accept the proposed value otherwise go to step 1.

Values from an exponential distribution can be simulated using the inversion method.

2. In the rejection sampling algorithm, the total acceptance rate is $1/C$, one can therefore use the ratio of accepted samples to provide an estimate for C.

Problem 2

Problem 2a)

The MCMC algorithm uses a random walk proposal. This can be recognised by the fact that the proposal is symmetric with regard to the current and proposed value and therefore the ratio of proposals is always 1 (and the lo-ratio is 0)

Problem 3

The correct alternatives are 1 and 3. For the other two alternative the distribution of the linear predictor η_i is not Gaussian.

Problem 4

The first statement is used in a frequentist setting. The interpretation is as follows: One defines two statistics $\hat{\theta}_{\text{lower}}$ and $\hat{\theta}_{\text{upper}}$ as function of a sample $x_1, \dots, x_n$ such that the stochastic interval $[\hat{\theta}_{\text{lower}}, \hat{\theta}_{\text{upper}}]$ contains the unknown paramter θ with 95% probability. The value of these statistics computed on the observed data is $\hat{\theta}_{\text{lower}} = 2$, $\hat{\theta}_{\text{upper}} = 2.5$. The statemen "the probability that θ lies in the [2,2.5] interval is 95%" therefore, does not make sense.

The second statement is used in Bayesian setting. Its interpretation is as follows: With some prior on θ, $\pi(\theta)$ (not specified in the question), the posterior probability that θ is in the interval [2.3, 2.5] is 95%, i.e $\int_2^{2.5} \pi(\theta|x_1, \dots, x_n)d\theta = 0.95$

Problem 5

Problem 5b)

The complete data likelihood is:

$$L(\theta; y_1, \ldots, y_n, x_1, \ldots, x_n) = \prod_{i:x_i=0} \frac{1}{\sqrt{2\pi}} \exp\{-\frac{1}{2}y_i^2\} \prod i : x_i = 1 \exp\{-y_i\} \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i}$$

and the log-likelihood

$$\begin{aligned} l(\theta; y_1, \ldots, y_n, x_1, \ldots, x_n) &= -\sum_{i:x_i=0} \frac{1}{2}y_i^2 - \sum_{i:x_i=1} y_i + \sum_{i=1}^{n} I(x_i = 1)\log(\theta) + \sum I(x_i = 0)\log(1-\theta) \\ &= \sum_{i=1}^{n} I(x_i = 0)\{-\frac{1}{2}y_i^2 + \log(1-\theta)\} + \sum_{i=1}^{n} I(x_i = 1)\{-y_i + \log\theta\} \end{aligned}$$

The $Q(\theta|\theta')$ is defined as

$$Q(\theta|\theta') = E(l(\theta; y_1, \ldots, y_n, x_1, \ldots, x_n)|y_1, \ldots, y_n, \theta')$$

which in out case becomes:

$$\begin{aligned} Q(\theta|\theta') = \sum_{i=1}^{n} E[I(x_i = 0)|y_1, \ldots, y_n, \theta']\{-\frac{1}{2}y_i^2 + \log\theta\} + \\ \sum_{i=1}^{n} E[I(x_i = 1)|y_1, \ldots, y_n, \theta']\{-y_i + \log(1-\theta)\} \end{aligned}$$

so we what we need is to compute $E[I(x_i = 1)|y_1, \ldots, y_n, \theta']$ and $E[I(x_i = 0)|y_1, \ldots, y_n, \theta']$.

We get:

$$\begin{aligned} E[I(x_i = 1)|y_1, \ldots, y_n, \theta'] = P(x_i = 1|y_1, \ldots, y_n, \theta') &= \frac{f(y_i|x_i = 1, \theta')P(x_i = 1|\theta')}{f(y_i|x_i = 1, \theta')P(x_i = 1|\theta') + f(y_i|x_i = 0, \theta')P(x_i = 0|\theta')} \\ &= \frac{\exp(y_i)\theta'}{\exp(y_i)\theta' + \frac{1}{\sqrt{2\pi}}\exp(-0.5y_i^2)(1-\theta')} = w_i \end{aligned}$$

and

$$\begin{aligned} E[I(x_i = 0)|y_1, \ldots, y_n, \theta'] = P(x_i = 0|y_1, \ldots, y_n, \theta') &= \frac{f(y_i|x_i = 0, \theta')P(x_i = 1|\theta')}{f(y_i|x_i = 1, \theta')P(x_i = 1|\theta') + f(y_i|x_i = 0, \theta')P(x_i = 0|\theta')} \\ &= \frac{\frac{1}{\sqrt{2\pi}}\exp(-0.5y_i^2)(1-\theta')}{\exp(y_i)\theta' + \frac{1}{\sqrt{2\pi}}\exp(-0.5y_i^2)(1-\theta')} = 1 - w_i \end{aligned}$$

so that the $Q(\theta|\theta')$ becomes:

$$\begin{aligned} Q(\theta|\theta') &= \sum_{i=1}^{n}(1-w_i)(-\frac{1}{2}y_i^2 + log(1-\theta)) + w_i(-y_i + \log\theta) \\ &= K + \log(1-\theta)\sum(1-w_i) + \log\theta\sum w_i \end{aligned}$$

where K is a quantity independent of θ.

To maximize wrt θ we have to compute the first derivative and set it to 0:

$$\frac{dQ(\theta|\theta')}{d\theta} = \frac{-1}{1-\theta}\sum(1-w_i) + \frac{1}{\theta}\sum w_i = 0$$

which gives $\theta = \frac{1}{n}\sum w_i$.

Problem 5c)

The standard deviation of the maximum likelihood estimator can be estimated by usinf the following algorithm:

For $b = 1, \dots, B$ *do*

1. Draw a boorstrap sample \$y_1^{b},..., y_n^{b} \$ by resampling the oberved data \$y_1,..., y_n \$
2. Use the EM algorithm to compute the ML estimator based on the bootstrap sample. Indicate the ML estimator as $\hat{\theta}_b^*$

Estimate the SD of $\hat{\theta}$ as:

$$\hat{\text{SD}}(\hat{\theta}) = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}(\hat{\theta}_b^* - \bar{\hat{\theta}}^*)}$$

where

$$\bar{\hat{\theta}}^*) = \frac{1}{B-1}\sum_{b=1}^{B}\hat{\theta}_b^*$$

Problem 6

The mode is found by setting the first derivative equal to 0

$$\frac{df(x)}{dx} = C\exp\{k_1 x - k_2 e^x\}(k_1 - k_2 e^x) = 0$$

The solution is found at $x = \log k_1 - \log k_2$.

Using the Taylor expansion around x_0 we get that

$$h(x) \approx a + bx - \frac{1}{2}cx^2$$

where $b = h'(x_0) - x_0 f''(x_0)$ and $c = -f''(x_0)$. We want to expand around the mode so in our case $x_0 = \log k_1 - \log k_2$.

After some algebra we get that:

$$b = k_1 \log\frac{k1}{k2} \text{ and } c = k_1$$

Out Gaussian approximation built around the mode of $f(x)$ is then

$$\tilde{f}(x) \propto \exp(-\frac{1}{2}k_1 x^2 + k_1 \log\frac{k1}{k2}x)$$

which is the canonical form of a Gaussian distribution with variance $1/k_1$ and mean $\log\frac{k_1}{k_2}$.

To find an approximation to the normalizing constant C one can notice that at the mode x_0 we have that

$$f(x_0) \approx \tilde{f}(x_0)$$

We have that

$$f(x_0) = C\exp\{k_1 \log k_1 + k_1 \log k_2 - k_1\}$$

and

$$\tilde{f}(x_0) = \sqrt{\frac{k_1}{2\pi}}$$

So our estimate of the constant becomes

$$\hat{C} = \sqrt{\frac{k_1}{2\pi}} \exp\{-k_1(\log k_1 - \log k_2 - 1)\} \approx 0.003$$

Problem 7

To apply the Gibbs sampling algorithm we need to find the two full conditional distribution from the joint

$$g(x_1, x_2) = C \;\exp(-x_1^2 x_2 + x_1 \log x_2)$$

We have that:

$$f(x_1|x_2) \propto \exp(-x_2 x_1^2 + log x_2 x_1$$

Here we recognise the core of a Gaussian distribution with variance $\sigma^2(x_2) = 1/(2x_2)$ and mean $\mu(x_2) = \frac{\log x_2}{2x_2}$. Moreover we have that

$$f(x_2|x_1) \propto \exp(-x_1^2 x_2 + x_1 \log x_2) = x_2^{x_1} \exp(-x_1^2 x_2)$$

and this is the core of a Gamma distribution with parameters $a(x_1) = x_1 + 1$ and $b(x_1) = x_1^2$. The Gibbs algorith is then:

Set a starting value x_1^0 and x_2^0

For $i = 1, \ldots, N$ *repeat*

1. Sample $x_1^i \sim \mathcal{N}(\mu(x_2^{i-1}, sigma^2(x_2^{i-1})))$
2. Sample $x_2^i \sim \text{Gamma}(a(x_1^i), b(x_1^i))$