Group 10 – HA2 Report

SF2955: Computer Intensive Methods in Mathematical Statistics

David Haapanen Noah Eriksson Tewolde Berhane

May 21, 2025

1 Problem 1: Bayesian Analysis of Coal Mine Disasters

1.1 a) Marginal Posterior Computations

We seek to compute the marginal posterior distributions of $f(\theta \mid \boldsymbol{\lambda}, \boldsymbol{t}, \boldsymbol{\tau})$, $f(\boldsymbol{\lambda} \mid \theta, \boldsymbol{t}, \boldsymbol{\tau})$, and $f(\boldsymbol{t} \mid \theta, \boldsymbol{\lambda}, \boldsymbol{\tau})$ up to an normalizing constant. Here, $\boldsymbol{\tau} = (\tau_1, ..., \tau_n)$ denotes the observed time points of the $n = 191$ disasters, $\boldsymbol{t} = (t_1, ..., t_{d+1})$ contains the fixed end points coupled with all the breakpoints. Furthermore $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_n)$ contains the intensities λ_i for the interval $[t_i, t_{i+1})$

Marginal posterior of θ

Utilizing *Bayes' Theorem* one can rewrite the marginal posterior as proportional to the likelihood times the prior,

$$f(\theta \mid \boldsymbol{\lambda}, \boldsymbol{t}, \boldsymbol{\tau}) \propto f(\boldsymbol{\tau} \mid \boldsymbol{t}, \boldsymbol{\lambda}, \theta) \cdot f(\boldsymbol{t}, \boldsymbol{\lambda}, \theta)$$

where the likelihood can be simplified as $f(\boldsymbol{\tau} \mid \boldsymbol{t}, \boldsymbol{\lambda}, \theta) = f(\boldsymbol{\tau} \mid \boldsymbol{t}, \boldsymbol{\lambda})$. Furthermore, the breakpoints $\boldsymbol{t}$ are independent of the intensities $\boldsymbol{\lambda}$ and θ such that

$$f(\boldsymbol{t}, \boldsymbol{\lambda}, \theta) = f(\boldsymbol{t}) \cdot f(\boldsymbol{\lambda}, \theta) = f(\boldsymbol{t}) \cdot f(\boldsymbol{\lambda} \mid \theta) \cdot f(\theta).$$

Thus we can consider the expression of

$$f(\theta \mid \boldsymbol{\lambda}, \boldsymbol{t}, \boldsymbol{\tau}) \propto f(\boldsymbol{\tau} \mid \boldsymbol{t}, \boldsymbol{\lambda}) \cdot f(\boldsymbol{t}) \cdot f(\boldsymbol{\lambda} \mid \theta) \cdot f(\theta)$$

Here only $f(\boldsymbol{\lambda} \mid \theta)$ and $f(\theta)$ are the terms that will depend on θ since the prior, $f(\boldsymbol{\tau} \mid \boldsymbol{t}, \boldsymbol{\lambda})$, for given values of $\boldsymbol{\lambda}$ and $\boldsymbol{t}$ doesn't vary w.r.t. θ. Thus discarding these terms we have that

$$f(\theta \mid \boldsymbol{\lambda}, \boldsymbol{t}, \boldsymbol{\tau}) \propto f(\boldsymbol{\lambda} \mid \theta) \cdot f(\theta)$$

Knowing that the prior for each intensity is $\lambda_i \mid \theta \sim \Gamma(2, \theta)$ and $\theta \sim \Gamma(2, \vartheta)$, whilst all λ_i are independent we finally have that

$$f(\theta \mid \boldsymbol{\lambda}, \boldsymbol{t}, \boldsymbol{\tau}) \propto f(\boldsymbol{\lambda} \mid \theta) \cdot f(\theta) \propto \theta^{(2d+2)-1} \exp\left\{-\theta \cdot \left(\vartheta + \sum_{i=1}^{d} \lambda_i\right)\right\}$$

which upon inspection is a $\Gamma(2d + 2, \vartheta + \sum_{i=1}^{d} \lambda_i)$ distribution.

Marginal posterior of $\boldsymbol{\lambda}$

As described from before the marginal posterior can be rewritten with the likelihood times the prior using *Bayes' Theorem*, furthermore the simplifications mentioned from the previous section still holds. Thus, it holds that

$$f(\boldsymbol{\lambda} \mid \theta, \boldsymbol{t}, \boldsymbol{\tau}) \propto f(\boldsymbol{\tau} \mid \boldsymbol{t}, \boldsymbol{\lambda}) \cdot f(\boldsymbol{t}) \cdot f(\boldsymbol{\lambda} \mid \theta) \cdot f(\theta)$$

where only $f(\boldsymbol{\tau} \mid \boldsymbol{t}, \boldsymbol{\lambda})$ and $f(\boldsymbol{\lambda} \mid \theta)$ will depend on $\boldsymbol{\lambda}$. Hence we can conclude that

$$f(\boldsymbol{\lambda} \mid \theta, \boldsymbol{t}, \boldsymbol{\tau}) \propto f(\boldsymbol{\tau} \mid \boldsymbol{t}, \boldsymbol{\lambda}) \cdot f(\boldsymbol{\lambda} \mid \theta)$$

Each λ_i is independent from each other and follow with a $\sim \Gamma(2, \theta)$ distribution. This coupled with that $f(\boldsymbol{\tau} \mid \boldsymbol{\lambda}, \boldsymbol{t}) \propto \prod_{i=1}^{d} \lambda_i^{n_i(\boldsymbol{\tau})} \exp\left(-\lambda_i(t_{i+1} - t_i)\right)$ gives the final expression of

$$f(\boldsymbol{\lambda} \mid \theta, \boldsymbol{t}, \boldsymbol{\tau}) \propto f(\boldsymbol{\tau} \mid \boldsymbol{t}, \boldsymbol{\lambda}) \cdot f(\boldsymbol{\lambda} \mid \theta) \propto \prod_{i=1}^{d} \lambda_i^{n_i(\boldsymbol{\tau})+1} \exp\left(-\lambda_i(\theta + t_{i+1} - t_i)\right)$$

The independence between the λ_i terms further yields that $\lambda_i \sim \Gamma(n_i(\boldsymbol{\tau}) + 2, \theta + t_{i+1} - t_i)$

Marginal posterior of $\boldsymbol{t}$

Following the same steps from before we know that

$$f(\boldsymbol{t} \mid \theta, \boldsymbol{\lambda}, \boldsymbol{\tau}) \propto f(\boldsymbol{\tau} \mid \boldsymbol{t}, \boldsymbol{\lambda}) \cdot f(\boldsymbol{t}) \cdot f(\boldsymbol{\lambda} \mid \theta) \cdot f(\theta)$$

where only the terms $f(\boldsymbol{\tau} \mid \boldsymbol{t}, \boldsymbol{\lambda})$ and $f(\boldsymbol{t})$ depend on $\boldsymbol{t}$ such that

$$\begin{aligned} f(\boldsymbol{t} \mid \theta, \boldsymbol{\lambda}, \boldsymbol{\tau}) &\propto f(\boldsymbol{\tau} \mid \boldsymbol{t}, \boldsymbol{\lambda}) \cdot f(\boldsymbol{t}) \propto \left(\prod_{i=1}^{d} \lambda_i^{n_i(\boldsymbol{\tau})} \exp\left(-\lambda_i(t_{i+1} - t_i)\right)\right) \left(\prod_{i=1}^{d} (t_{i+1} - t_i)\right) \\ &\propto \{\lambda_i \text{ are considered as constans}\} \propto \exp\left(-\sum_{i=1}^{d} \lambda_i (t_{i+1} - t_i)\right) \cdot \prod_{i=1}^{d} (t_{i+1} - t_i) \end{aligned}$$

for $t_1 < t_2 < \cdots < t_d < t_{d+1}$. This has however no standard or clear distribution.

1.2 b) Hybrid MCMC Algorithm

Gibbs sampling (GS) was implemented combined with a Metropolis-Hastings (MH) step with the chosen proposal distribution of a random walk (provided in the assignment). A pseudo-code for the full algorithm, with $\vartheta = 1$, $\rho = 1$ and $d - 1$ breakpoints, can be seen below

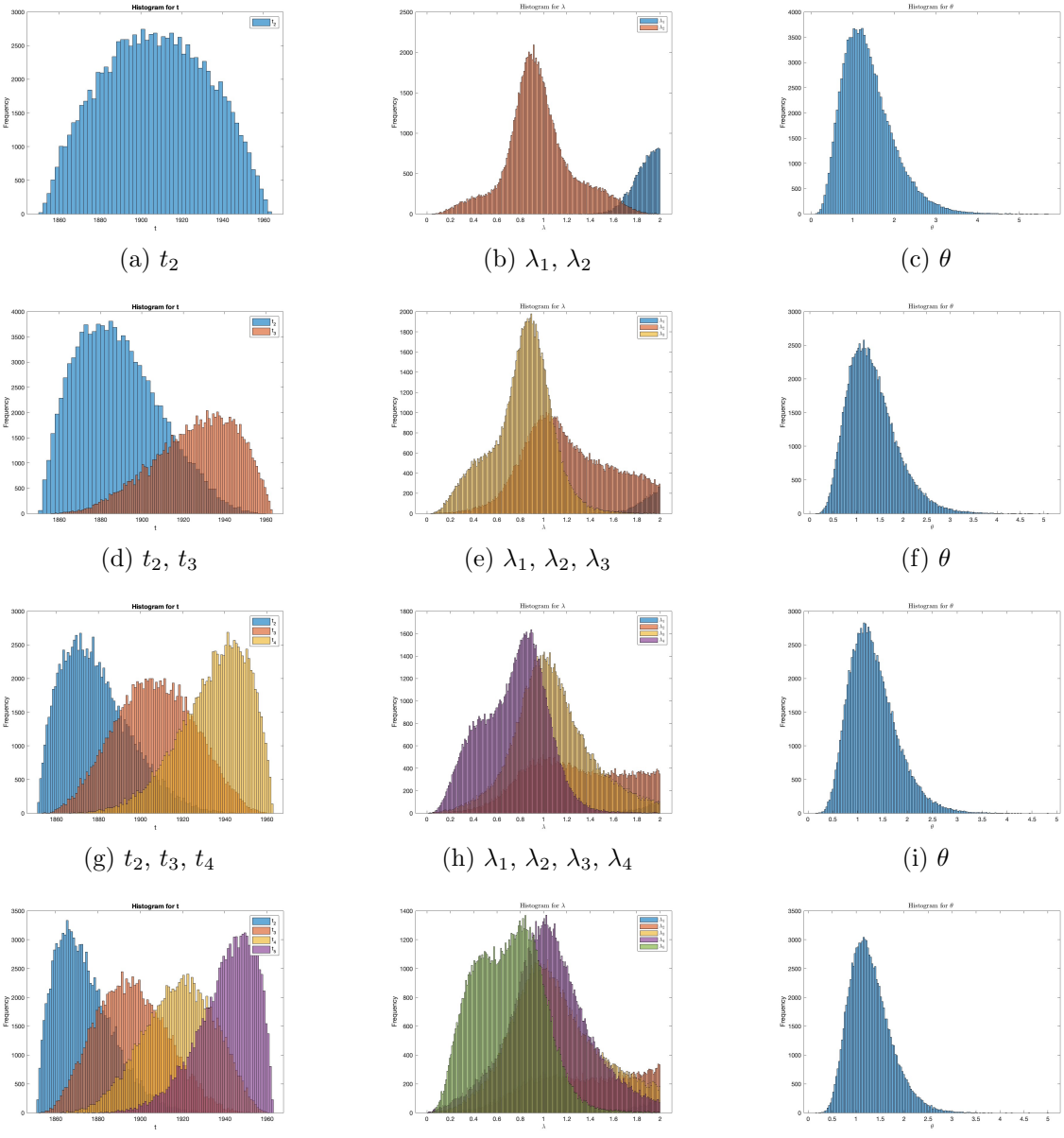
Algorithm 1 Hybrid MCMC

Parameters: $N = 10^5$, $\vartheta = 1$, $\rho = 1$
for $i = 1$ to $N - 1$ **do**
  Draw $\theta^{i+1} \sim f(\theta \mid \boldsymbol{\lambda}^i, \boldsymbol{t}^i, \boldsymbol{\tau})$
  Draw $\boldsymbol{\lambda}^{i+1} \sim f(\boldsymbol{\lambda} \mid \theta^{i+1}, \boldsymbol{t}^i, \boldsymbol{\tau})$
  for $k = 2$ to d **do**
    Draw $\varepsilon \sim \text{Unif}(-R, R)$ with $R = \rho(t_{k-1}^i - t_{k+1}^i)$
    Propose $t_k^* = t_k^i + \varepsilon$
    if $t_{k-1} < t_k^* < t_{k+1}$ **then**
      Compute $\alpha = \min\left(1, \frac{f(t_k^*)}{f(t_k)} = \frac{(t_k^* - t_{k-1})(t_{k+1} - t_k^*)}{(t_k - t_{k-1})(t_{k+1} - t_k)}\right)$
      if Unif(0,1) $\leq \alpha$ **then**
        $t_k^{i+1} \leftarrow t_k^*$
      else
        $t_k^{i+1} \leftarrow t_k^i$
      end if
    end if
  end for
end for

Histograms of the distributions from sampled θ, $\boldsymbol{\lambda}$ and $\boldsymbol{t}$ were furthermore obtained when letting the number of breakpoints be 1, 2, 3 and 4 which can be seen in the different rows in fig. 1.

1.3 c) Behavior of MCMC with varying amount of breakpoints

The distribution of θ, illustrated in fig. 1, doesn't change for the varying amount of breakpoints having mean ≈ 1.33. The distributions of $\boldsymbol{t}$ are also similar in their shape but have different means for the varying amount of break points. This is not strange since one introduces new equidistant break points for the different cases. Furthermore the varying histograms of the $\boldsymbol{\lambda}$ distributions, adding more breakpoints produces increasingly overlapping intensity distributions. Overlapping peaks for the different intensities indicate that they may have the same intensity hence introducing extra break points may fail to capture any genuinely distinct shifts in disaster rates, and are thus redundant. This is clearly seen in fig. 1k as λ_3 and λ_4 have distributions that overlap. On the other hand having at least one break point can be strongly argued for as the peaks between the distribution of λ_1 and λ_2 are very distinct as illustrated in fig. 1b.



(a) t_2 (b) λ_1, λ_2 (c) θ

(d) t_2, t_3 (e) λ_1, λ_2, λ_3 (f) θ

(g) t_2, t_3, t_4 (h) λ_1, λ_2, λ_3, λ_4 (i) θ

(j) t_2, t_3, t_4, t_5 (k) λ_1, λ_2, λ_3, λ_4, λ_5 (l) θ

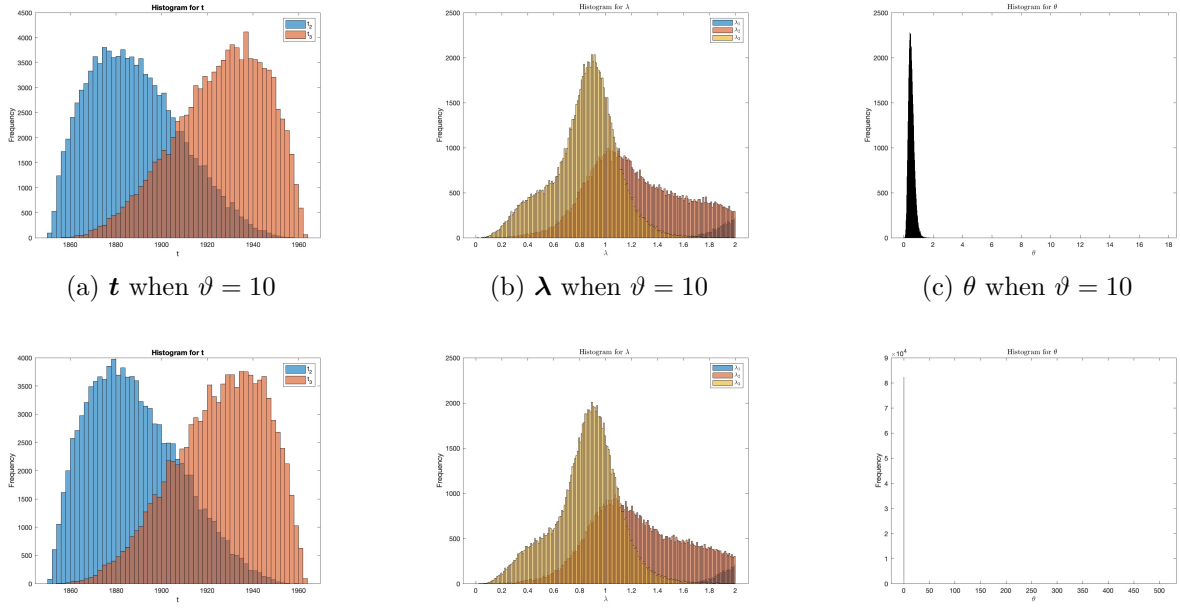
Figure 1: Distributions of θ, $\boldsymbol{\lambda}$, and $\boldsymbol{t}$ using 1, 2, 3 and 4 breakpoints, with $\vartheta = 1$ and $\rho = 1$.

1.4 d) Sensitivity to choice of ϑ

Histograms of $\boldsymbol{t}$, $\boldsymbol{\lambda}$ and θ for $\rho = 1$, $\vartheta = 1$, and 2 breakpoints can be seen as the second row of fig. 1. To investigate the sensitivity to the hyperparameter ϑ, the MCMC algorithm was applied with $\vartheta = 10$ and $\vartheta = 100$. The resulting histograms are displayed in fig. 2. When comparing these to the corresponding baseline histograms mentioned in the begining of this section, we observe minor differences in the posterior distributions of $\boldsymbol{t}$ and $\boldsymbol{\lambda}$ whilst the variation in distribution of θ changes drastically. Since $\theta \sim \Gamma(2, \vartheta)$ the change in distribution of θ is expected. The result indicate that the distributions of $\boldsymbol{t}$ and $\boldsymbol{\lambda}$ are non-sensitive to changes in θ, and therefore ϑ. This further implies that the prior distribution $f(\theta)$ doesn't affect the distributions of $\boldsymbol{t}$ and $\boldsymbol{\lambda}$.

1.5 e) Sensitivity to choice of ρ

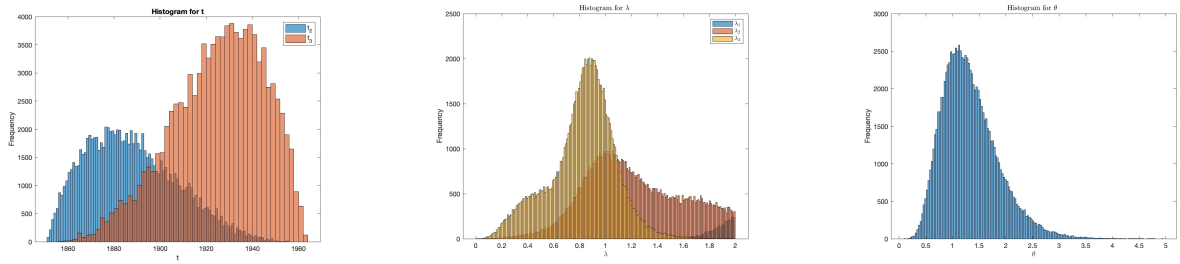
Similarly from the previous question the MCMC algorithm was applied for $\vartheta = 1$ and 2 breakpoints when letting $\rho = 2$ and $\rho = 3$ to produce the histograms seen in fig. 3. In comparison to the second row of fig. 1 one can see the magnitude of t_2 and t_3 varies for the different cases while the prior distribution of $\boldsymbol{\lambda}$ and θ is unchanged. This suggests that only the prior distribution of $\boldsymbol{t}$ is sensitive to the value of ρ. Furthermore the mixing of the MH step was observed by editing the algorithm in section 1.2 such that the acceptance rate in the MH step could be evaluated. When varying ρ one could see that the acceptance rate would also change. An indication for good mixing is when the acceptance rate is $\approx 30\%$ which could be continuously observed for when $\rho = 1.25$ (even for the varying amount of breakpoints). Thus the mixing of the MH step is sensitive to ρ which aligns with the sensitivity of the prior distribution of $\boldsymbol{t}$. That is the draws are made using MH sampling hence candidates are either more likely to be rejected or accepted dependent on ρ, which correspondingly will affect the histograms of $\boldsymbol{t}$.



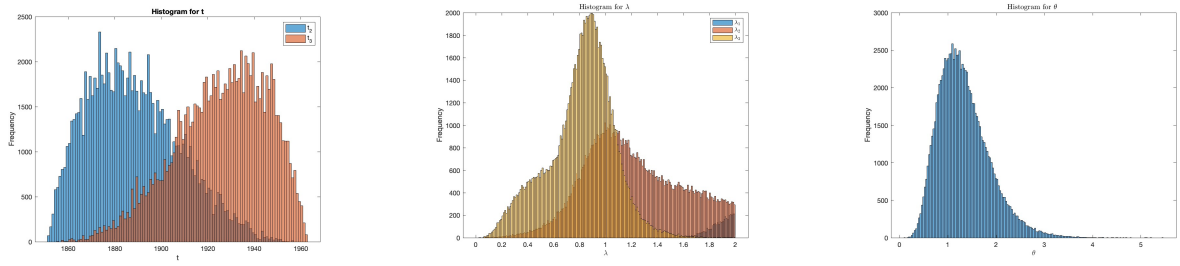
(a) $\boldsymbol{t}$ when $\vartheta = 10$ (b) $\boldsymbol{\lambda}$ when $\vartheta = 10$ (c) θ when $\vartheta = 10$

(d) $\boldsymbol{t}$ when $\vartheta = 100$ (e) $\boldsymbol{\lambda}$ when $\vartheta = 100$ (f) θ when $\vartheta = 100$

Figure 2: Distributions of $\boldsymbol{t}$, $\boldsymbol{\lambda}$, and θ for $\rho = 1$, two breakpoints, and varying ϑ.



(a) $\boldsymbol{t}$ when $\rho = 2$ (b) $\boldsymbol{\lambda}$ when $\rho = 2$ (c) θ when $\rho = 2$



(d) $\boldsymbol{t}$ when $\rho = 3$ (e) $\boldsymbol{\lambda}$ when $\rho = 3$ (f) θ when $\rho = 3$

Figure 3: Distributions of $\boldsymbol{t}$, $\boldsymbol{\lambda}$, and θ for $\vartheta = 1$ and 2 breakpoints under varying ρ.

2 Problem 2: Sampling from a circle-shaped Posterior using Hamiltonian Monte Carlo

2.1 a) Posterior $\ln f(\boldsymbol{\theta} \mid \boldsymbol{y})$ and its gradient $\nabla_{\boldsymbol{\theta}} \ln f(\boldsymbol{\theta} \mid \boldsymbol{y})$

Using Baye's theorem we have that $f(\boldsymbol{\theta} \mid \boldsymbol{y}) \propto f(\boldsymbol{y} \mid \boldsymbol{\theta}) \cdot f(\boldsymbol{\theta})$ which upon taking the natural logarithm results in the expression of

$$\ln f(\boldsymbol{\theta} \mid \boldsymbol{y}) \propto \ln f(\boldsymbol{y} \mid \boldsymbol{\theta}) + \ln f(\boldsymbol{\theta})$$

Furthermore using that each $y_i \mid \boldsymbol{\theta} \sim \mathcal{N}(\theta_1^2 + \theta_2^2 = \|\boldsymbol{\theta}\|^2, \sigma^2)$ and that each y_i are i.i.d one can get the following expressions

$$f(\boldsymbol{y} \mid \boldsymbol{\theta}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - \|\boldsymbol{\theta}\|^2)^2}{2\sigma^2}\right) \Rightarrow \ln f(\boldsymbol{y} \mid \boldsymbol{\theta}) \propto -\sum_{i=1}^{n} \frac{(y_i - \|\boldsymbol{\theta}\|^2)^2}{2\sigma^2}$$

$$\frac{\partial}{\partial \theta_j} \ln f(\boldsymbol{y} \mid \boldsymbol{\theta}) = 2 \sum_{i=1}^{n} \frac{(y_i - \|\boldsymbol{\theta}\|^2)}{\sigma^2} \cdot \theta_j$$

It is also given that $f(\boldsymbol{\theta}) \sim \mathcal{N}(\mathbf{0}, \Sigma)$, where using that Σ^{-1} is symmetric yields in the following expressions

$$f(\boldsymbol{\theta}) = \frac{1}{2\pi|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}\boldsymbol{\theta}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}\right) \Rightarrow \ln f(\boldsymbol{\theta}) \propto -\frac{1}{2}\boldsymbol{\theta}^\top \Sigma^{-1}\boldsymbol{\theta}$$

$$\frac{\partial}{\partial \boldsymbol{\theta}} \ln f(\boldsymbol{\theta}) \propto \frac{\partial}{\partial \boldsymbol{\theta}} \left(-\frac{1}{2}\boldsymbol{\theta}^\top \Sigma^{-1}\boldsymbol{\theta}\right) \propto -\frac{2}{2} \cdot \Sigma^{-1}\boldsymbol{\theta} \propto -\Sigma^{-1}\boldsymbol{\theta}$$

Thus the resulting expression of $\ln f(\boldsymbol{\theta} \mid \boldsymbol{y})$ and its gradient becomes

$$\ln f(\boldsymbol{\theta} \mid \boldsymbol{y}) \propto \ln f(\boldsymbol{y} \mid \boldsymbol{\theta}) + \ln f(\boldsymbol{\theta}) \propto -\sum_{i=1}^{n} \frac{(y_i - \|\boldsymbol{\theta}\|^2)^2}{2\sigma^2} - \frac{1}{2}\boldsymbol{\theta}^\top \Sigma^{-1}\boldsymbol{\theta}$$

$$\nabla_{\boldsymbol{\theta}} \ln f(\boldsymbol{\theta} \mid \boldsymbol{y}) \propto \frac{2}{\sigma^2} \sum_{i=1}^{n} (y_i - \|\boldsymbol{\theta}\|^2) \cdot \boldsymbol{\theta} - \Sigma^{-1}\boldsymbol{\theta}$$

2.2 b) Hamiltonian Monte Carlo Algorithm Design

Using the previously determined posterior of $\ln f(\boldsymbol{\theta} \mid \boldsymbol{y})$ and $\nabla_{\boldsymbol{\theta}} \ln f(\boldsymbol{\theta} \mid \boldsymbol{y})$ an algorithm for sampling from $f(\boldsymbol{\theta} \mid \boldsymbol{y})$ was designed using the Hamiltonian Monte Carlo (HMC) method. Similar to the MH algorithm an auxiliary proposal is used for drawing candidate samples. The HMC proposal is based on Hamiltonian dynamics, involving the gradient of the desired distribution to propose more informed moves leading to faster mixing. The core of HMC is the *Leapfrog* integrator, which is a volume-preserving involution, yielding proposals that are accepted with high probability. The integrator solves Hamilton's equations numerically to propose new $(\boldsymbol{\theta}, v)$ pairs that explore the target distribution efficiently. Given an initial momentum v a *Leapfrog* integrator utilized for time length $T = \varepsilon \cdot L$, using L iterations of time-step ε can be described as:

$$
\begin{aligned}
v_{m+1/2} &\leftarrow v_m - \tfrac{\varepsilon}{2} \nabla U(\boldsymbol{\theta}_m), \\
\boldsymbol{\theta}_{m+1} &\leftarrow \boldsymbol{\theta_m} + \varepsilon\, v_{m+1/2}, \\
v_{m+1} &\leftarrow v_{m+1/2} - \tfrac{\varepsilon}{2} \nabla U(\boldsymbol{\theta}_{m+1}),
\end{aligned}
$$

for $m \in \{0, 1, ..., L-1\}$ before applying the sign-shift operator $\sigma(\boldsymbol{\theta}, v) = (\boldsymbol{\theta}, -v)$. The newly computed $\boldsymbol{\theta}$ is then either accepted or rejected in a Metropolis step. A pseudo code of the HMC algorithm can be seen below.

Algorithm 2 Hamiltonian Monte Carlo (HMC)

Parameters: ε, L, N, observed data $\boldsymbol{y}$, σ, $\boldsymbol{\Sigma}$, initial value $\boldsymbol{\theta}^{(0)}$

for $i = 2$ to N **do**
  Sample momentum: $\mathbf{v}^{(i)} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{2\times 2})$
  Set $\boldsymbol{\theta}_{\text{current}} \leftarrow \boldsymbol{\theta}^{(i)}$
  Perform leapfrog integration using step size ε and L iterations
    $[\boldsymbol{\theta}^*, \mathbf{v}^*] \leftarrow \texttt{Leapfrog}(\boldsymbol{\theta}_{\text{current}}, \mathbf{v}^{(i)}, \epsilon, L)$
  Compute: $\alpha = \min\left(1, \exp\left(\ln f(\boldsymbol{\theta^*} \mid \boldsymbol{y}) - \ln f(\boldsymbol{\theta}_{current} \mid \boldsymbol{y}) - \frac{1}{2}\|\mathbf{v}^*\|^2 + \frac{1}{2}\|\mathbf{v}^{(i)}\|^2\right)\right)$
  Sample $U \sim \text{Unif}(0, 1)$
  if $U \leq \alpha$ **then**
    $\boldsymbol{\theta}^{(i+1)} \leftarrow \boldsymbol{\theta}^*$
  else
    $\boldsymbol{\theta}^{(i+1)} \leftarrow \boldsymbol{\theta}^{(i)}$
  end if
end for
Output: Samples $\{\boldsymbol{\theta}^{(i)}\}_{i=1}^{N}$

2.3 c) Comparison between MH and HMC sampling

Using the outline of the HMC algorithm provided in section 2.2 with $N = 100\,000$ the distribution of $f(\boldsymbol{\theta} \mid \boldsymbol{y})$ was sampled using the given parameters σ and $\boldsymbol{\Sigma}$ combined with the set values for ε and L. The values for these were set by using various combinations and assessing autocorrelation plots of θ_1 and θ_2 and acceptance rates of the HMC algorithm.

- Smaller values of L could lead to higher acceptance rates but continuously showed autocorrelation plots with slower decay. Since the autocorrelation plots showed more correlation between samples the higher acceptance rate was not seen as beneficial.
- Larger values of ε made the assessed distribution of $f(\boldsymbol{\theta} \mid \boldsymbol{y})$ look less like the provided plot from the assignment. The autocorrelation functions would also display values that were further away from zero.
- Smaller values of ε would make the computational time of the algorithm longer but show

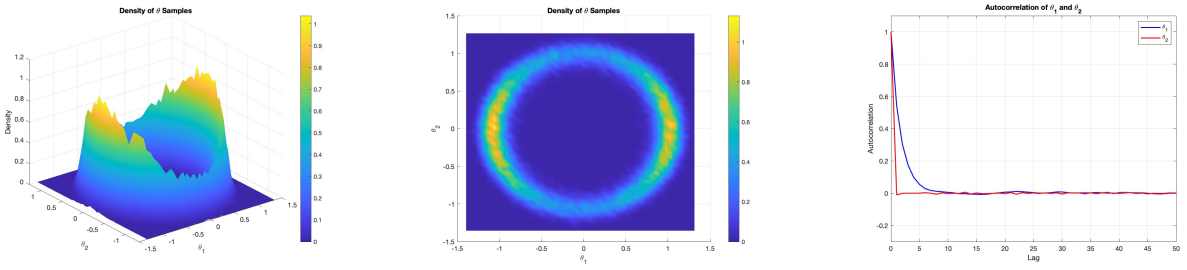
less correlation between samples in the autocorrelation plots.

Based on the observation the values were set to $\varepsilon = 0.09$ and $L = 25$, where it was found that these values continuously led to a acceptance rate of $\approx 77\%$ while the samples of $\boldsymbol{\theta}$ seem to have sufficiently low correlation between them. The sampled distribution of $f(\boldsymbol{\theta} \mid \boldsymbol{y})$ and the autocorrelation plots of θ_1 and θ_2 can be seen in fig. 4.

$N = 100\,000$ samples from the distribution of $f(\boldsymbol{\theta} \mid \boldsymbol{y})$ was also retrieved using a MH algorithm with a bivariate random walk with the candidate having the distribution of $\boldsymbol{\theta}^* \sim \mathcal{N}(\boldsymbol{\theta}, \zeta^2 \boldsymbol{I}_{2\times 2})$). The tuning of ζ was done by trying various values and assessing the acceptance rate of the algorithm.

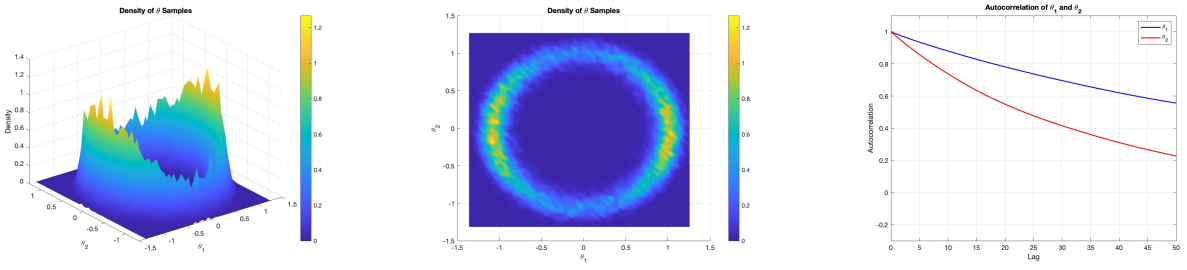
- The acceptance rate was $\approx 32\%$ around $\zeta = 0.35$
- The autocorrelation plots showed very strong correlation between samples for the varied values of ζ, but slightly lower around the chosen $\zeta = 0.35$.
- The sampled distribution of $f(\boldsymbol{\theta} \mid \boldsymbol{y})$ was generally very unstable and not smooth looking compared to the plot provided in the assignment.

Autocorrelation plots of θ_1 and θ_2 and the sampled distribution of $f(\boldsymbol{\theta} \mid \boldsymbol{y})$ using the chosen parameter value $\zeta = 0.35$ can be seen in fig. 5



(a) Surface plot of $f(\boldsymbol{\theta} \mid \boldsymbol{y})$ (b) 2D histogram of $f(\boldsymbol{\theta} \mid \boldsymbol{y})$ (c) Autocorrelation of θ_1 and θ_2

Figure 4: Distribution and autocorrelation plots using HMC algorithm with $\varepsilon = 0.09$, $L = 25$.



(a) Surface plot of $f(\boldsymbol{\theta} \mid \boldsymbol{y})$ (b) 2D histogram of $f(\boldsymbol{\theta} \mid \boldsymbol{y})$ (c) Autocorrelation of θ_1 and θ_2

Figure 5: Distribution and autocorrelation plots using MH algorithm with $\zeta = 0.35$.

After drawing $N = 100{,}000$ samples using the MH and HMC algorithm to estimate the distribution of $f(\boldsymbol{\theta} \mid \boldsymbol{y})$, it is clear that HMC provides a more accurate estimate of the target distribution. The HMC algorithm samples seen in fig. 4b show a circular shape which has a more even edge while the samples from the MH are mostly drawn from the right or left focal points seen in fig. 5b. One can also see that the distribution in fig. 4a resembles the provided posterior distribution from the assignment more in comparison to fig. 5a. The difference is also supported by the autocorrelation plots. In fig. 4c, the autocorrelation decreases rapidly, indicating that the HMC samples are less correlated and mix more efficiently. In contrast, fig. 5c shows a much slower decay in autocorrelation, meaning the MH samples are more dependent.