HW2 solutions

Problem 1 (a)

$$F(x) = \int_{-\infty}^{x} f(x)\mathrm{d}\,x = \begin{cases} \frac{1}{2}e^{x}, & x < 0 \\ 1 - \frac{1}{2}e^{-x}, & x \geq 0 \end{cases},$$

So we can generate the samples of $f(x)$ using

$$x = F^{-1}(u) = \begin{cases} \ln(2u), & u < \frac{1}{2} \\ -\ln(2-2u), & u \geq \frac{1}{2} \end{cases}, \quad u \sim \text{Uniform}(0,1).$$

(b) Consider $Cf(x)$ as candidate distributions for the envelop function. To make sure it is valid, we need

$$Cf(x) \geq \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}.$$

We can choose

$$C = \max_{x} \frac{2}{\sqrt{2\pi}}e^{-\frac{x^2}{2}+|x|} = \sqrt{\frac{2e}{\pi}}.$$

The rejection sampling works as follows:

1. Draw a sample x from the Laplace density $f(x)$.
2. Generate $u \sim U[0,1]$.
3. If $u \leq \frac{\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}}{Cf(x)}$, then accept x as a new sample; otherwise, discard it.
4. return to step 1.

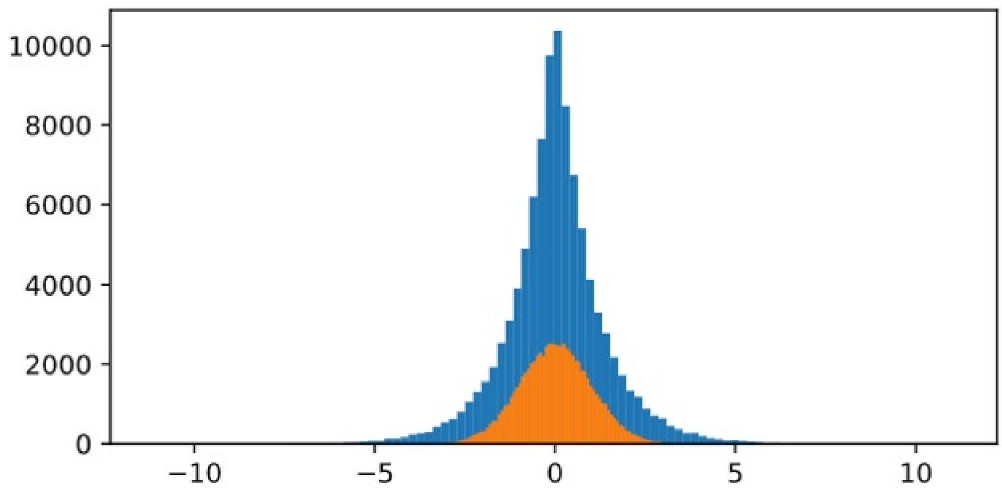

Figure 1: Sample frequency of standard Laplace distribution and the accepted samples.

(c) We can not simulate Laplace random variables using rejection sampling with a multiple of the standard normal density as the envelop. For the reason that $f(x)/(\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}})$ is unbounded.

Problem 2 (a)

$$\mathbb{E}\hat{\pi}(x) = \mathbb{E}_{y\sim q}\frac{\pi(x,y)}{q(y)} = \int \frac{\pi(x,y)}{q(y)}q(y)\mathrm{d}\,y = \pi(x)$$

(b) Consider the augmented Markov chain with the state space $(x, z_1, \cdots, z_K)$, the transfer probability is

$$p(x', z_1', \cdots, z_K'|x, z_1, \cdots, z_K) = Q(x'|x)\alpha(x'|x)\prod_{i=1}^{K} q(z_i')$$

We have

$$\begin{aligned}\frac{p(x, z_1, \cdots, z_K|x', z_1', \cdots, z_K')}{p(x', z_1', \cdots, z_K'|x, z_1, \cdots, z_K)} &= \frac{Q(x|x')\alpha(x|x')\prod_{i=1}^{K} q(z_i)}{Q(x'|x)\alpha(x'|x)\prod_{i=1}^{K} q(z_i')}\\ &= \frac{\min\{\hat{\pi}(x')Q(x|x'), \hat{\pi}(x)Q(x'|x)\}\hat{\pi}(x)\prod_{i=1}^{K} q(z_i)}{\min\{\hat{\pi}(x)Q(x'|x), \hat{\pi}(x')Q(x|x')\}\hat{\pi}(x')\prod_{i=1}^{K} q(z_i')}\\ &= \frac{\hat{\pi}(x)\prod_{i=1}^{K} q(z_i)}{\hat{\pi}(x')\prod_{i=1}^{K} q(z_i')}.\end{aligned}$$

So the stationary distribution of the augmented Markov chain is $\hat{\pi}(x)\prod_{i=1}^{K} q(z_i)$. For the original Markov chain, the marginal distribution of x is its stationary distribution, i.e.

$$\int \hat{\pi}(x)\prod_{i=1}^{K} q(z_i)\mathrm{d}\,z_1\cdots\mathrm{d}\,z_K = \frac{1}{K}\sum_{i=1}^{K}\int\int \pi(x, z_i)\mathrm{d}\,z_i\mathrm{d}\,z_{-i} = \frac{1}{K}\sum_{i=1}^{K}\pi(x) = \pi(x).$$

Problem 3 (a) The joint posterior prabability of (β, σ, z) is

$$p(\beta, \sigma^2, z|y, x) \propto p(\beta)p(\sigma^2)\prod_{i=1}^{N} p(z_i|\beta, \sigma^2)p(y_i|z_i)$$

where $p(\beta)$ and $p(\sigma^2)$ is the prior distribtion, $p(z_i|\beta, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{(z_i - x_i\beta)^2}{2\sigma^2})$ and $p(y_i|z_i) = 1_{y_i=1, z_i\geq 0} + 1_{y_i=0, z_i<0}$. The update rule of Gibbs sampling is

$$z_i^{(t+1)} \sim N(x_i\beta^{(t)}, (\sigma^{(t)})^2) \text{ truncated by } z \geq 0 \text{ if } y_i = 1, z < 0 \text{ if } y_i = 0$$

$$\beta^{(t+1)} \sim N\left(\frac{100\sum x_i z_i^{(t+1)}}{(\sigma^{(t)})^2 + 100\sum x_i^2}, \frac{100(\sigma^{(t)})^2}{(\sigma^{(t)})^2 + 100\sum x_i^2}\right)$$

$$(\sigma^{(t+1)})^2 \sim \mathrm{Inv}-\chi^2\left(3+n, \frac{3 + \sum(z_i - x_i\beta^{(t+1)})^2}{3+n}\right)$$

The result of Gibbs sampling starting at $(25, 25)$ is shown below.

(b) The transition probability is $Q(\beta', (\sigma^2)'|\beta, \sigma^2) = \frac{1}{\beta}\exp(-\frac{1}{\beta}\beta')\frac{1}{\sigma^2}\exp(-\frac{1}{\sigma^2}(\sigma^2)')$. As the joint posterior distribution of (β, σ^2) is $p(\beta, \sigma^2|y, x) \propto \prod_{i=1}^{n}\Phi\left(\frac{x_i\beta}{\sigma}\right)^{y_i}\Phi\left(-\frac{x_i\beta}{\sigma}\right)^{1-y_i}\cdot\exp(-\frac{\beta^2}{200})\cdot\exp(-\frac{3}{2\sigma^2})\cdot\sigma^{-5}$. Thus the acceptence probability is $A((\beta, \sigma^2), (s\beta, s\sigma^2))) = \min(1, a((\beta, \sigma^2), (s\beta, s\sigma^2)))$

where

$$
\begin{aligned}
&a((\beta,\sigma^2),(s_1\beta,s_2\sigma^2)) \\
=&\frac{Q(\beta|s_1\beta)Q(\sigma^2|s_2\sigma^2)p(s_1\beta,s_2\sigma^2|y,x)}{Q(s_1\beta|\beta)Q(s_2\sigma^2|\sigma^2)p(\beta,\sigma^2|y,x)} \\
=&\frac{1}{s_1 s_2}\exp\left(s_1+s_2-\frac{1}{s_1}-\frac{1}{s_2}\right)s_2^{-\frac{5}{2}}\frac{\prod_{i=1}^{n}\Phi\left(\frac{x_i\beta s_1}{\sigma\sqrt{s_2}}\right)^{y_i}\Phi\left(-\frac{x_i\beta s_1}{\sigma\sqrt{s_2}}\right)^{1-y_i}}{\prod_{i=1}^{n}\Phi\left(\frac{x_i\beta}{\sigma}\right)^{y_i}\Phi\left(-\frac{x_i\beta}{\sigma}\right)^{1-y_i}}\exp\left(-\frac{(s_1^2-1)\beta^2}{200}-\frac{3}{2\sigma^2}(\frac{1}{s_2}-1)\right)
\end{aligned}
$$

Thus the MH algorithm is constructed as follows.

1. sample $s_1, s_2 \sim Exp(1)$;
2. calculate $A((\beta,\sigma^2),(s_1\beta,s_2\sigma^2))) = \min(1, a((\beta,\sigma^2),(s_1\beta,s_2\sigma^2)))$;
3. sample $r \sim U([0,1])$;
4. accept $(s_1\beta, s_2\sigma^2)$ if $r < A((\beta,\sigma^2),(s_1\beta,s_2\sigma^2)))$ else reject.

We can find the sampling procedure converges faster with MH steps inserted.

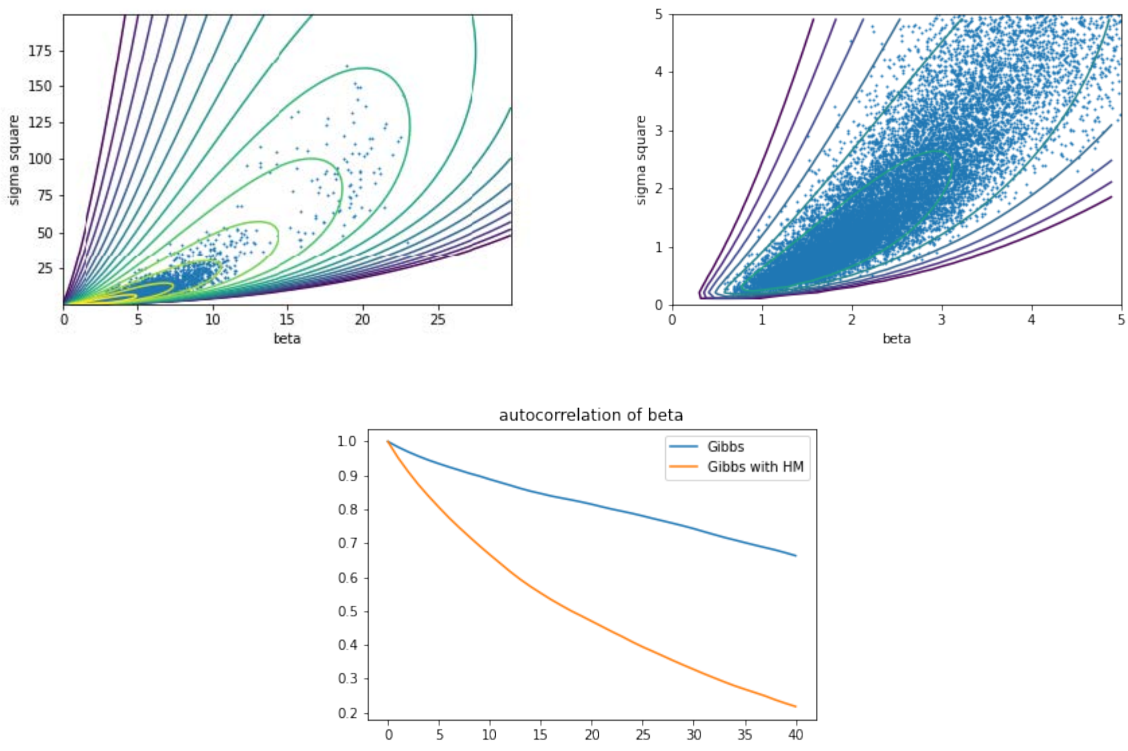

Figure 2: Results of Gibbs sampling without (left) and with (right) MH step

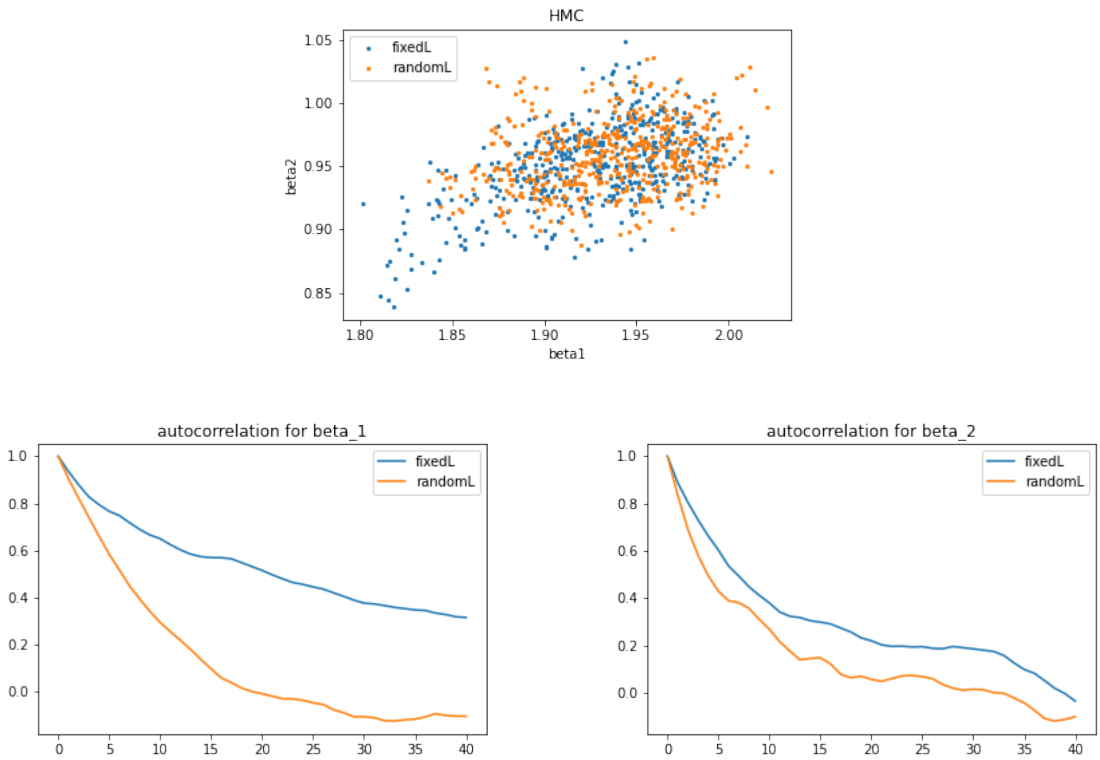

Figure 3: HMC with fixed L and random L: scatter plot and autocorrelation plot

Problem 4 (a) The log posterior distribution of β is

$$\begin{aligned}
\log p(\beta|y) &= C + \log\left[p(\beta)\prod p(y_i|\beta)\right] \\
&= C - \frac{1}{2}\beta^T\beta - \sum \log(1+\exp(-x_i^T\beta)) - \sum(x_i^T\beta)(1-y_i)
\end{aligned}$$

and the gradient of $\log p(\beta|y)$ w.r.t. β is

$$\nabla \log p(\beta|y) = -\beta + \sum x_i^T y - \sum \frac{x_i^T}{1+\exp(-x_i^T\beta)}.$$

Thus the leapfrog update rule for HMC is

$$\begin{aligned}
r(t+\frac{\varepsilon}{2}) &= r(t) + \frac{\varepsilon}{2}\nabla \log p(\beta^{(t)}|y) \\
\beta(t+\varepsilon) &= \beta(t) + r(t+\frac{\varepsilon}{2})\varepsilon \\
r(t+\varepsilon) &= r(t+\frac{\varepsilon}{2}) + \frac{\varepsilon}{2}\nabla \log p(\beta^{(t+\epsilon)}|y)
\end{aligned}$$

The acceptance probability is $A(\beta', r'|\beta, r) = \min(1, a(\beta', r'|\beta, r))$ where

$$\begin{aligned}
a(\beta', r'|\beta, r) &= \exp(-H(\beta', r') + H(\beta, r)) \\
&= \exp(\log p(\beta'|y) - \log p(\beta|y) + \frac{1}{2}r^T r - \frac{1}{2}r'^T r')
\end{aligned}$$

We can use, for example, $L = 5$ as the fixed number of leapfrog steps, and $U(1, 2L-1)$ as the random number of leapfrog steps. We can find that HMC with random L converges faster.

(b) Suppose N, n is the full sample size and batch size. We can compute the stochastic gradient $\tilde{g}(\beta)$ of $\log p(\beta|y)$ w.r.t β by

$$\tilde{g}(\beta) = -\beta + \frac{N}{n}\left[\sum_{1:n} x_i^T y - \sum_{1:n} \frac{x_i^T}{1+\exp(-x_i^T\beta)}\right].$$

The update scheme of SGMCMC algorithms are as follows.

- SGLD:
$$\beta^{(t+1)} = \beta^{(t)} + \frac{\varepsilon_t}{2}\tilde{g}(\beta) + \sqrt{\varepsilon_t}\eta_t, \quad \eta_t \sim N(0, I).$$
We can set $\varepsilon = 0.0001$.

- SGHMC:
$$\begin{aligned}\beta^{(t)} &= \beta^{(t-1)} + \varepsilon_t r^{(t-1)} \\ r^{(t)} &= r^{(t-1)} + \varepsilon_t \tilde{g}(\beta^{(t)}) - \varepsilon_t C\beta^{(t)} + \sqrt{2C\varepsilon_t}\eta_t, \quad \eta_t \sim N(0, I)\end{aligned}$$
We set $C = 1$ and $\varepsilon = 0.001$.

- SGNHT:
$$\text{Initialized } \beta^{(0)}, r^{(0)} \sim N(0, I), \xi^{(0)} = A$$
$$\begin{aligned}r^{(t)} &= r^{(t-1)} + \varepsilon_t \tilde{g}(\beta^{(t-1)}) - \varepsilon_t \xi^{(t-1)} r^{(t-1)} + \sqrt{2A\varepsilon_t}\eta, \quad \eta \sim N(0, I) \\ \beta^{(t)} &= \beta^{(t-1)} + \varepsilon_t r^{(t)} \\ \xi^{(t)} &= \xi^{(t-1)} + \varepsilon_t((r^{(t)})^T r^{(t)}/d - 1)\end{aligned}$$
$d = 2$ in this problem. For SGNHT, we set $A = 35, \varepsilon = 0.005$.

For all these three SGMCMC algorithms, KL divergence converges to zero as the number of iteration increase. In fact, after 10000 iterations, KL divergence is smaller than 0.05 for all these three algorithms.

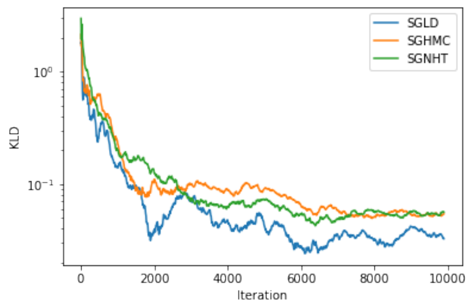

Figure 4: KL divergence between samples of different SGMCMC algorithms and the ground truth