Exercise 03

Wang, Haotian 421723

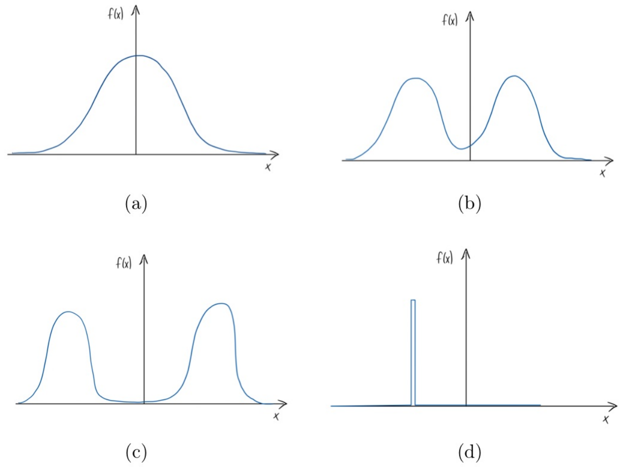
Li, Kehao, 461037

Wang, Yinghao 460777

Tao, Tingxu 453995

Exercise 1 **(Markov Chain Monte Carlo)** **25P**

Consider the following density functions:



(a). rejection sampling, because it has only one peak and can be approximated by a normal probability density function,

so $C = \sup_x \left\{ \frac{f(x)}{g(x)} \right\}$ won't be too large.

(b). Metropolis Hasting, because the pdf has 2 peaks,

and Metropolis Hasting is more efficient than rejection sampling.

(c). Rejection sampling. Metropolis Hasting is not suitable because $f(x)$ is close to 0 when $|x|$ is small,

so if the starting point is in the middle, it will accept until it reaches one of the peaks, and then it is more likely

to always reject, so it is hard to get to the other peak, which does not correspond to the original pdf $f(x)$

(d). Rejection Sampling, but it takes a lot of times to obtain a sample.

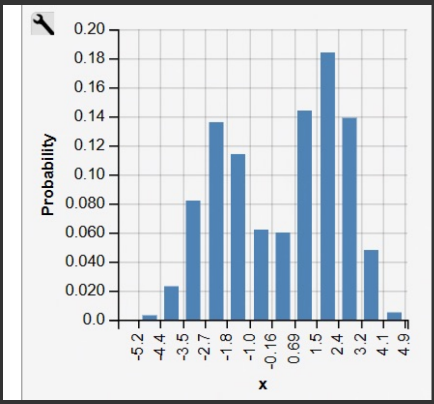
Metropolis Sampling is not suitable because $I = \{x \mid f(x) \gg 0\}$ is a narrow interval.

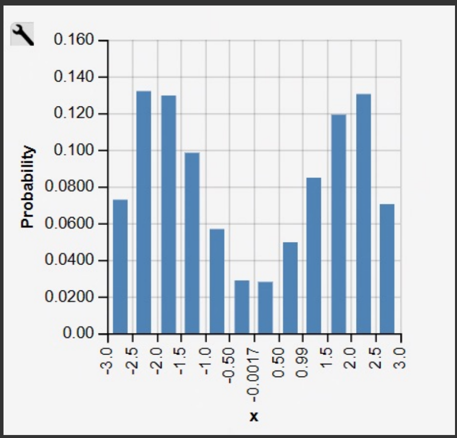
If starting point x_0 is far away from I, the chance of rejection is still high because $f(x)$ is almost

constant around x_0.

Ex 02:

(b).





(c). The larger σ^2 is, the more samples will be rejected, which leads to more same samples apearing next to each other in the sample sequence and less different samples.

For example, we tested for $n = 20000$ samples from $f(x)=\frac{1}{2}e^{-\frac{(x-2)^2}{2}}+\frac{1}{2}e^{-\frac{(x+2)^2}{2}}$

Let n_A denote the number of accepted samples.

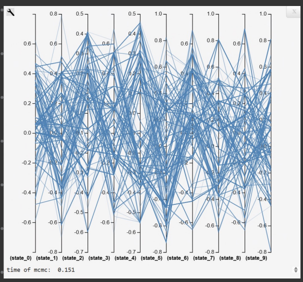
$\sigma^2 = 0.3$, $n_A = 18377$

$\sigma^2 = 10$, $n_A = 4757$

$\sigma^2 = 0.01$, $n_A = 19932$.

Ex 03.

(a).



(b). Metropolis Hasting: 0.141s

Rejection Sampling: 13.042s.

(c). $P = \frac{\frac{2\pi^{d/2}}{d \cdot \Gamma(\frac{d}{2})}}{1^d} = \frac{2\pi^{\frac{d}{2}}}{d \cdot \Gamma(\frac{d}{2})}$

$P|_{d=10} = 0.510$, $P|_{d=20} = 2.58 \times 10^{-3}$ $P|_{d=40} = 1.802 \times 10^{-10}$

Let X be a random variable that represents the number of samples to obtain an accepted sample.

$$P(X=k) = (1-p)^{k-1} p.$$

Therefore, $X \sim Geo(p)$, so $E(X) = \frac{1}{p}$

To get $n=1000$ accepted samples, the expeted total number of samples is

$$\frac{n}{p} = \frac{1000}{p} = \frac{1000}{2\pi^{\frac{d}{2}}} \cdot d \cdot \Gamma(\frac{d}{2}).$$

For $d=10$, $\frac{n}{p} = 1.96 \times 10^3$ order of magnitude: 10^3.

$d=20$, $\frac{n}{p} = 3.87 \times 10^5$ order of magnitude: 10^5.

$d=40$, $\frac{n}{p} = 5.55 \times 10^{12}$ order of magnitude: 10^{12}.

Exercise 4 (Detailed Balance Condition) **30P**

Let $D = (\Sigma, \sigma_0, \mathcal{P})$ be a Markov chain with $n = |\Sigma|$ states. We say that a Markov chain D satisfies the *detailed balance condition* for a distribution vector $\rho \in \mathbb{R}^n$ if for all states $\sigma_i, \sigma_j \in \Sigma$ it holds that

$$\rho(i) \cdot \mathcal{P}(i,j) = \rho(j) \cdot \mathcal{P}(j,i).$$

A Markov chain which satisfies the detailed balance condition for ρ is called *time reversible* with respect to ρ.

(a) [10P] Show that if the Markov chain D satisfies the detailed balance condition for ρ, then ρ is a stationary distribution of D.

(b) [10P] A Markov chain D can also be viewed as a sequence of random variables $X_0, X_1, X_2, \ldots$, where X_k corresponds to the k-th state of D.

Fix a Markov chain D with stationary distribution ρ.
Now consider a sequence of random variables of length m, i.e., $Y_0, Y_1, \ldots, Y_m$, and assume that for all states $\sigma_i \in \Sigma$ we have that $\Pr(Y_0 = i) = \rho(i)$.

1. Argue briefly that given Y_{k+1}, the state Y_k is independent of $Y_{k+2}, Y_{k+3}, \ldots, Y_m$ (thus the reverse sequence is Markovian).
2. Specify the transition probabilities $\mathcal{Q}$ for the reverse sequence, where $\mathcal{Q}(i,j)$ tells us the probability to go from state i at time $k+1$ to state j at time k (with $0 \le k < k+1 \le m$).

(c) [10P] Consider a time reversible Markov chain with stationary distribution ρ. Prove that $\mathcal{Q} = \mathcal{P}$ holds, hence when considering the transitions in reverse order the transition probabilities remain the same.

(a). The j-th component of the outcome probability $\underline{\rho}' = \underline{\rho} \cdot \underline{\underline{P}}$ is:

$$\rho'_j = (\underline{\rho} \cdot \underline{\underline{P}})_j = \sum_{i=1}^{n} \rho(i) \cdot P(j|i)$$

$$= \sum_{i=1}^{n} \rho(i) \cdot P(i,j)$$

$$= \sum_{i=1}^{n} \rho(j) \cdot P(j,i)$$

$$= \rho(j) \underbrace{\sum_{i=1}^{n} P(i|j)}_{=1}$$

$$= \rho(j)$$

$$= \rho_j \quad \text{for } 1 \le j \le n.$$

Hence, $\rho' = \rho$, so ρ is a stationary distribution of D.

(b).

1. For given Y_k, Y_{k+1} is only dependent of Y_k

so if Y_{k+1} is given, the probability distribution of Y_k can be computed without the information of $Y_{k+2}, \dots, Y_m$. Hence the state Y_k is independent of $Y_{k+2}, \dots, Y_m$.

2. $Q(i,j) = Pr[Y_k = j \mid Y_{k+1} = i]$

$$= \frac{Pr[Y_{k+1} = i \mid Y_k = j] \cdot Pr(Y_k = j)}{\sum_{l=1}^{n} Pr[Y_{k+1} = i \mid Y_k = l] \cdot Pr(Y_k = l)}$$

Since ρ is stationary distribution and $Pr(Y_0 = i) = \rho(i)$,

then $Pr(Y_k = i) = Pr(Y_0 = i) = \rho(i)$

And hence,

$$Q(i,j) = \frac{Pr[Y_{k+1} = i \mid Y_k = j] \cdot Pr(Y_k = j)}{\sum_{l=1}^{n} Pr[Y_{k+1} = i \mid Y_k = l] \cdot Pr(Y_k = l)}$$

$$= \frac{P(j,i) \cdot \rho(j)}{\sum_{l=1}^{n} P(l,i) \cdot \rho(l)}$$

(c). $$Q(i,j) = \frac{P(j,i)\, \rho(j)}{\sum_{l=1}^{n} P(l,i) \cdot \rho(l)}$$

$$= \frac{P(i,j) \cdot \rho(i)}{\sum_{l=1}^{n} P(i,l) \cdot \rho(i)}$$

$$= P(i,j) \cdot \frac{\rho(i)}{\rho(i) \cdot \underbrace{\sum_{l=1}^{n} P(i,l)}_{=1}}$$

$= P(i,j)$ for all $1 \le i \le n$ and $1 \le j \le n$