Computational Methods (16:540:540)
Homework 2: Solution

1. (Monte Carlo Integration)

(a) Develop Matlab code to approximate the mean of a Beta(α, β) distribution with $\alpha = 3$ and $\beta = 4$ using Monte Carlo integration. You can use the Matlab function betarnd to draw samples from a Beta distribution. You can compare your answer with the analytic solution: $\alpha/(\alpha, \beta)$.

(b) Similarly, approximate the variance of a Gamma(a, b) distribution with $a = 1.5$ and $b = 4$ by Monte Carlo integration. The Matlab command gamrnd allows you to sample from this distribution. Your approximation should get close to the theoretically derived answer.

Solution: Monte Carlo integration approach can be used to approximate the integration. In Monte Carlo integration approach,
1, gerneate n random samples $x_1, x_2, ..., x_n$ from target distribution.
2, evaluate $\frac{1}{n}\sum_{i=1}^{n} x_i$ and $\frac{1}{n}\sum_{i=1}^{n}(x_i - \frac{1}{n}\sum_{i=1}^{n} x_i)^2$.

(a) The table below shows the average sample mean with n random sample in 25 tests, and the percentage difference between the true mean $\mu = \frac{\alpha}{\alpha+\beta}$. As the number of random sample n increase, the approximated mean is more and more close to the analytic solution (the true mean).

n	10	100	200	500	1000	10000	100000	1000000
mean	0.41461	0.42656	0.43172	0.42667	0.42902	0.42899	0.42863	0.42859
$\frac{\lvert\bar{x}-\mu\rvert}{\mu}$	0.08354	0.03188	0.02260	0.01737	0.00793	0.00449	0.00117	0.00030

The figures below shows the estimated mean of Beta(3,4) and the error rate with different sample sizes.

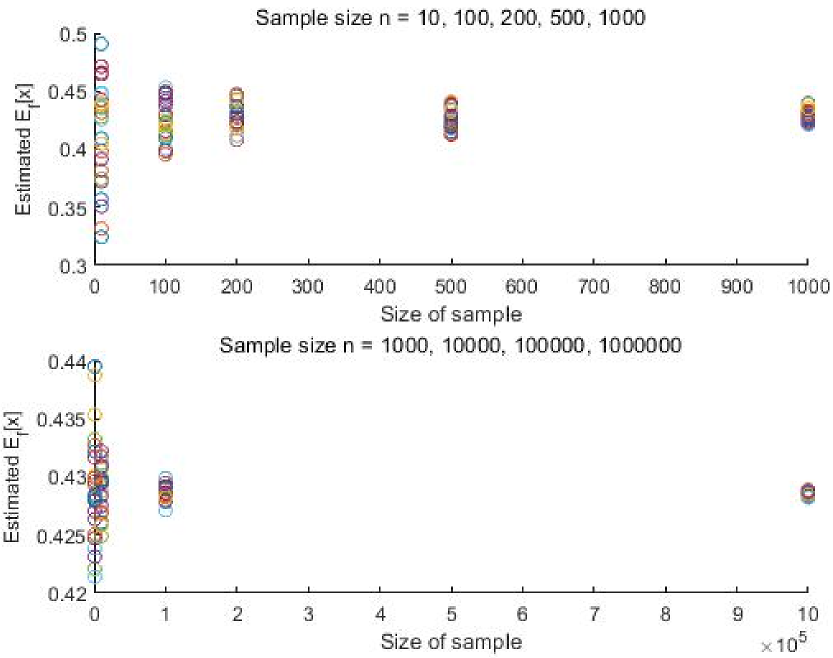


Figure 1: The estimated mean of Beta(3,4) in different sample sizes.

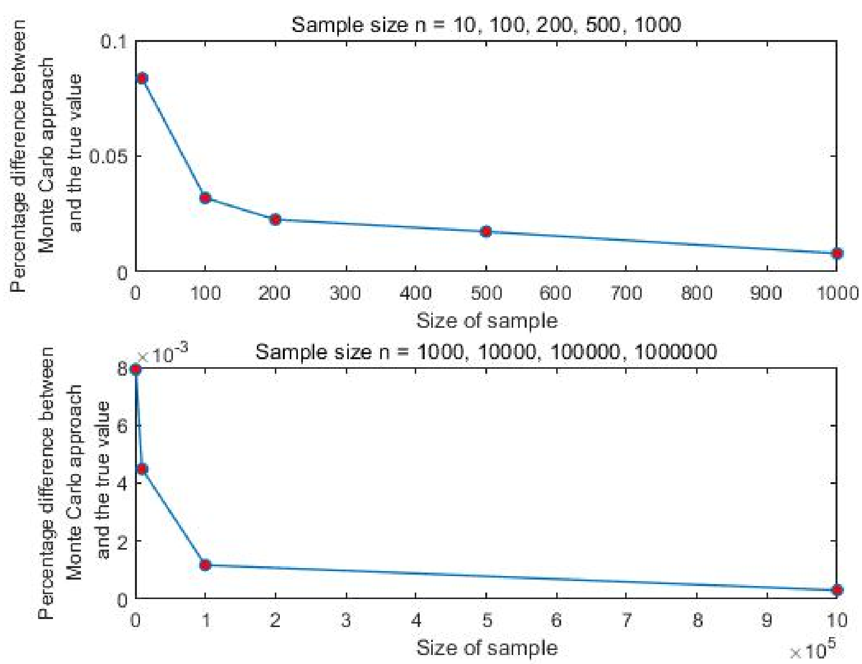


Figure 2: The error rate of the Monte Carlo approach with different sample sizes.

(b) Similarly, using Monte Carlo integration the approximation of the variance of a Gamma(a, b) distribution with $a = 1.5$ and $b = 4$ can be calculated as $\frac{1}{n}\sum_{i=1}^{n}(x_i - \frac{1}{n}\sum_{i=1}^{n} x_i)^2$. The theoretically derived answer is $\sigma^2 = ab^2 = 24$. The table below shows the average sample variance with n random sample in 25 tests, and the percentage difference between the true mean $\sigma^2 = a \cdot b^2 = 24$. As the number of random sample n increase, the approximated mean is more and more close to the analytic solution (the true mean).

n	10	100	200	500	1000	10000	100000	1000000
variance	31.0861	25.8221	25.3468	25.5525	24.4491	24.2567	23.9898	23.9964
$\frac{s-\sigma^2}{\sigma^2}$	0.72445	0.21547	0.17572	0.09530	0.07184	0.02342	0.00494	0.00216

The figures below shows the estimated mean of Gamma(3,4) and the error rate with different sample sizes.

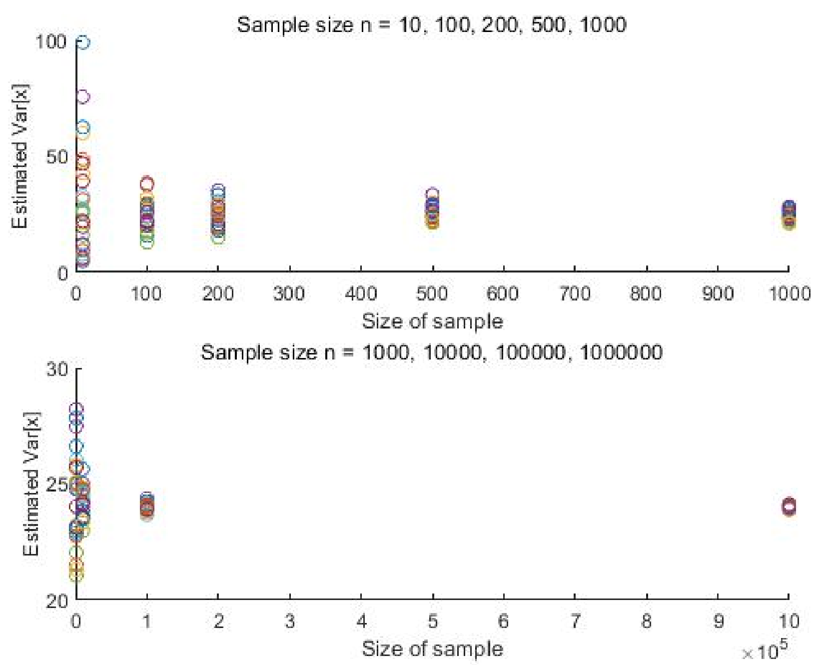


Figure 3: The estimated variance of Gamma(3,4) in different sample sizes.

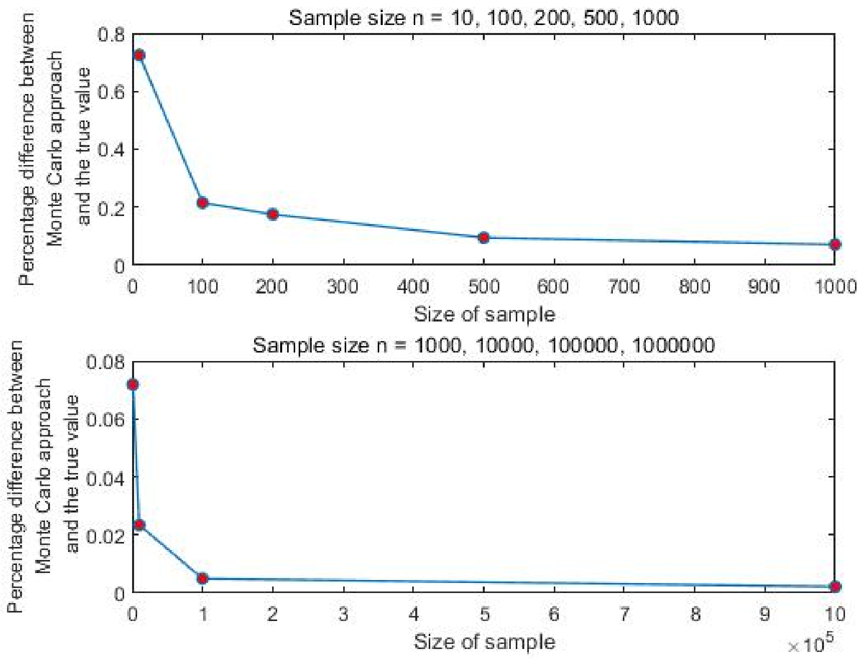


Figure 4: The error rate of the Monte Carlo approach with different sample sizes.

2. Suppose we wish to generate random samples from the Cauchy distribution using importance sampling and $MCMC$. The probability density of the Cauchy is given by:

$$f(\theta) = \frac{1}{\pi(1+\theta^2)}$$

We will use the Normal distribution as the proposal distribution. Our proposals are generated from a Normal$(\theta^{(t)}, \sigma^2)$ distribution.

(a) Generate 500 samples using
 a. the importance sampling technique
 b. the Metropolis sampler
 c. using the M-H sampler

(b) Estimate the probability density function using any density estimation technique based on the above samples from each method.

(c) Show the sequence of samples of each chain.

Solution: Let $h(x) = x$ and $g(x) = \frac{1}{\sqrt{2\pi}\times\sigma} e^{-\frac{(x-\theta)^2}{2\times\sigma^2}}$. We have,

$$\begin{aligned}
E_f[h(x)] &= \int_{-\infty}^{+\infty} h(x)\cdot f(x)dx \\
&= \int_{-\infty}^{+\infty} h(x)\cdot \frac{f(x)}{g(x)}\cdot g(x)dx \\
&= \int_{-\infty}^{+\infty} x\cdot \frac{\frac{1}{\pi(1+x^2)}}{\frac{1}{\sqrt{2\pi}\times\sigma}e^{-\frac{(x-\theta)^2}{2\times\sigma^2}}}\cdot g(x)dx \\
&= \int_{-\infty}^{+\infty} x\cdot \frac{\frac{\sqrt{2\pi}\times\sigma}{\pi(1+x^2)}}{e^{-\frac{(x-\theta)^2}{2\times\sigma^2}}}\cdot g(x)dx \\
&= \int_{-\infty}^{+\infty} x\cdot \frac{\sqrt{2\pi}\times\sigma}{\pi(1+x^2)}\cdot e^{\frac{(x-\theta)^2}{2\times\sigma^2}}\cdot g(x)dx \\
&= \int_{-\infty}^{+\infty} \frac{\sqrt{2\pi}\sigma}{\pi}\frac{x}{(1+x^2)}\cdot e^{\frac{(x-\theta)^2}{2\sigma^2}}\cdot g(x)dx
\end{aligned}$$

(a) Generate 500 samples using
 a. Importance Sampling approach,

1, gerneate $n = 500$ random samples $x_1, x_2, ..., x_n$ from the proposed distribution $N(\theta, \sigma^2)$, we can take $\theta = 0, \sigma = 1$.
2, For $i = 1, 2, ..., 500$, $x_i = \frac{\sqrt{2\pi}1}{\pi} \frac{x_i}{(1+x_i^2)} \cdot e^{\frac{(x_i-0)^2}{2\cdot 1^2}}$.

b. The steps of Metropolis sampler
• step 0: Generate θ_0 from $U(-1, 1)$.
• step 1: Generate a sample θ_* from the proposed distribution $N(\theta_{t-1}, \sigma^2)$.
• step 2: Calculate the acceptance probability, $\alpha = min(1, \frac{p(\theta_*)}{p(\theta_{t-1})}) = min(1, \frac{1+\theta_{t-1}^2}{1+\theta_*^2})$
• step 3: Generate a u from $U(0, 1)$. If $u \leq \alpha$, accept the generated sample, otherwise, $\theta_* = \theta_{t-1}$
• step 4: Repeat step 1-3, until $t = T$ or 500 random samples are generated.

c. The steps of M-H sampler
• step 0: Generate θ_0 from $U(-1, 1)$.
• step 1: Generate a sample θ_* from the proposed distribution $N(\theta_{t-1}, \sigma^2)$.
• step 2: Calculate the acceptance probability,

$$\alpha = min(1, \frac{p(\theta_*)q(\theta_{t-1}|\theta_*)}{p(\theta_{t-1})q(\theta_*|\theta_{t-1})}) = min(1, \frac{(1+\theta_{t-1}^2)\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(\theta_{t-1}-\theta_*)^2}{2\sigma^2}}}{(1+\theta_*^2)\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(\theta_*-\theta_{t-1})^2}{2\sigma^2}}}) = min(1, \frac{(1+\theta_{t-1}^2)}{(1+\theta_*^2)})$$

• step 3: Generate a u from $U(0, 1)$. If $u \leq \alpha$, accept the generated sample, otherwise, $\theta_* = \theta_{t-1}$
• step 4: Repeat step 1-3, until $t = T$ or 500 random samples are generated.

Here, the result should be similar to Metropolis sampler, since the proposal distribution is normal distribution, which is symmetric.

(b) For a general Cauchy distribution with two parameters(x_0, γ). The PDF is given as

$$\frac{1}{\pi\gamma\left[1+\left(\frac{x-x_0}{\gamma}\right)^2\right]}$$

The likelihood function $L(x_1, x_2, ..., x_{500}; x_0, \gamma)$ is

$$L(x_1, x_2, ..., x_{500}; x_0, \gamma) = f(x_1; x_0, \gamma) \cdot f(x_2; x_0, \gamma) \cdot ... \cdot f(x_{500}; x_0, \gamma) = \prod_{i=1}^{500} f(x_i; x_0, \gamma)$$

$$= \frac{1}{(\pi\gamma)^{500}} \frac{1}{\prod_{i=1}^{500} \left[1 + \left(\frac{x - x_0}{\gamma}\right)^2\right]}$$

Take ln on both side, we have the logarithm of L is

$$l(x_1, x_2, ..., x_{500}; x_0, \gamma) = -500 ln(\pi\gamma) - \sum_{i=1}^{500} ln \left[1 + \left(\frac{x_i - x_0}{\gamma}\right)^2\right]$$

$$\frac{\partial l}{\partial x_0} = -\sum_{i=1}^{500} \frac{1}{1 + \left(\frac{x_i - x_0}{\gamma}\right)^2} \cdot \frac{-2(x_i - x_0)}{\gamma^2} = 0 \Rightarrow \sum_{i=1}^{500} \frac{(x_i - x_0)}{\gamma^2 + (x_i - x_0)^2} = 0$$

$$\frac{\partial l}{\partial \gamma} = -500\frac{1}{\gamma} - \sum_{i=1}^{500} \frac{1}{1 + \left(\frac{x_i - x_0}{\gamma}\right)^2} \cdot \frac{-2(x_i - x_0)^2}{\gamma^3} = 0 \Rightarrow \sum_{i=1}^{500} \frac{(x_i - x_0)^2}{\gamma^2 + (x_i - x_0)^2} = 250$$

Estimate the probability density function using any density estimation technique based on the above samples from each method.

	Target	Importance Sampling	Metropolis and M-H	Inverse Transform Sampling
γ	1	0.4150	0.9690	0.9870
x_0	0	-0.0080	-0.1050	0.0340

The table above shows the MLE from all three sampling methods. Based on the difference between the target distribution and the parameter obtained from generated samples, the Inverse Transform Sampling provides the most similar samples with target distribution. And Figure 5, Figure 6, and Figure 7 show the histogram of the samples from three methods and the estimated p.d.f. The samples from IS are more concerted than the desired density. And the samples from MCMC are affected by the initially proposed mean. It may take a few iterations until it is stable.

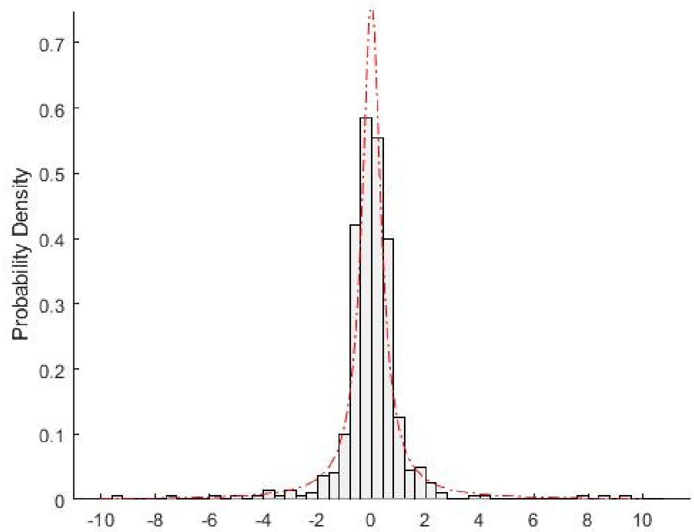


Figure 5: The samples generated from Importance Sampling method and the estimated p.d.f.

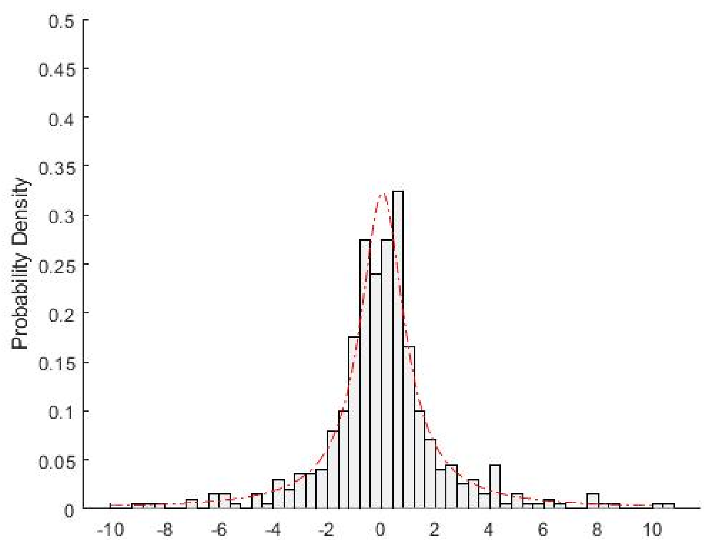


Figure 6: The samples generated from Inverse Transform Sampling method and the estimated p.d.f.

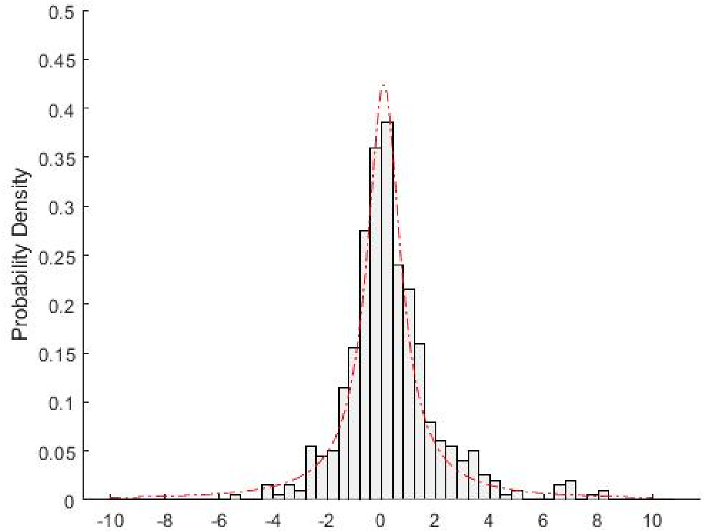


Figure 7: The samples generated from Metropolis method and the estimated p.d.f.

(c) Figure 8 shows the sequence of samples of each chain. The ITS provides the most volatile samples, and the samples from MCMC is the most smooth.

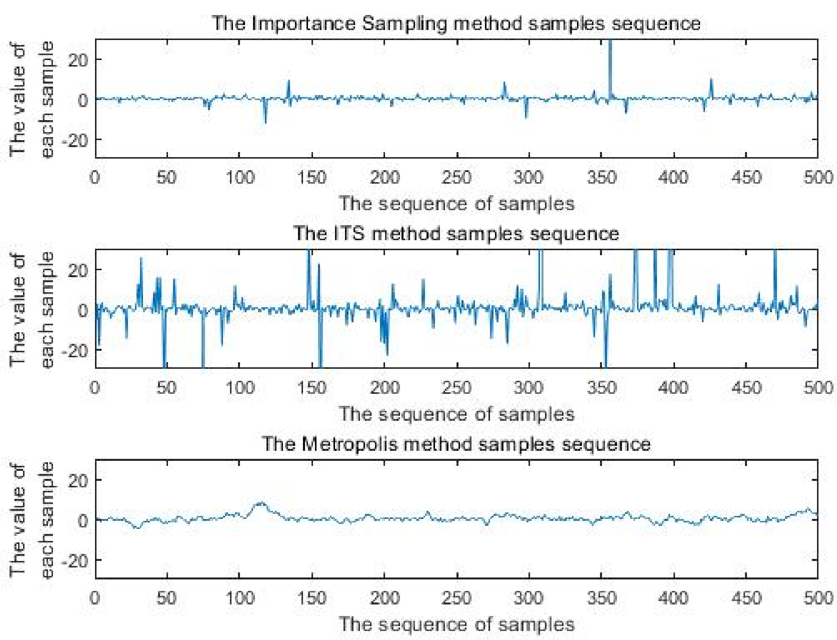


Figure 8: The sequence of samples of each chain.

3. Some additional questions for Problem #2.

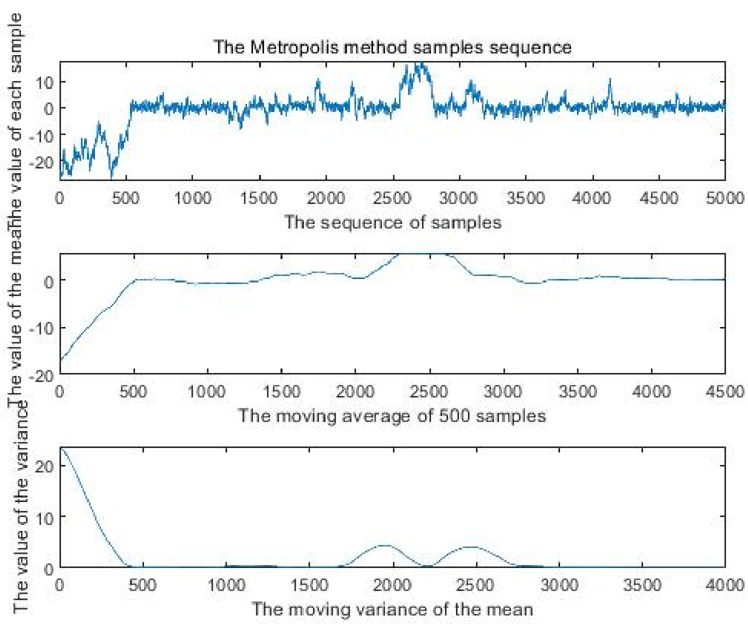
(a) Currently, in Problem #2 above, we take all states from the chain as samples to approximate the target distribution. Therefore, it also includes samples while the chain is still "burning in". Why is this not a good idea? Can you modify the code such that the effect of burn-in is removed?

(b) Explore the effect of different starting conditions. For example, what happens when we start the chain with $\theta = -30$?

(c) Calculate the proportion of samples that is accepted on average. Explore the effect of parameter σ on the average acceptance rate.

Solution:

(a) The solution above shows that there are almost no burn-in periods since the initial mean of the proposal distribution provides a good start sample, which is already in where the peak of p.d.f. However, if the start initial mean is set as -30, Figure 9 indicates the chain is unstable until about 700 samples are generated. If all the samples are taken into consideration, the samples generated in burn-in period will significantly impact the accuracy.



We can use the moving average method to calculate the mean of a chain of adjacent samples. Discard the samples until the mean becomes stable. We can also calculate the variance of the mean after moving average. Discard the samples generated at the beginning when the variance is away from zero.

(b) When the initial mean is set as $\theta = -30$, the chain will take a longer time to finish the "burn-in" period. The figure below shows that chain is stable after about 700 samples. A good starting solution could reduce the length of burn-in period.

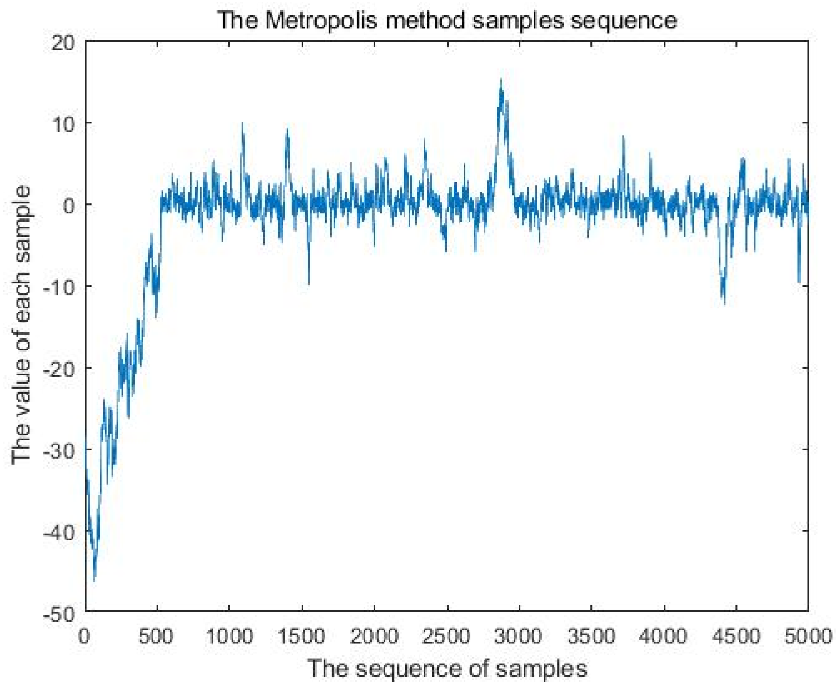


Figure 9: The sequence of samples from MCMC when the initial mean $\theta = -30$.

(c) For this Cauchy distribution, the acceptance rate will decline when the σ of proposal distribution increases. The p.d.f. of this Cauchy distribution is shaped as tall and narrow, the smaller σ means that it is similar to the target distribution, thus, the acceptance rate is high when σ is small.

σ	0.1	0.5	1.0	2.0	5.0	10	50	100
acceptance rate	0.976	0.862	0.755	0.617	0.405	0.264	0.087	0.050