Extra Credit: MCMC

Jeong Hyun Lee
STAT 5114 – Statistical Inference

April 29, 2025

Part 1

We provide the pseudocode for the Metropolis-Hastings sampler for $\mu, \phi \mid X$ in Algorithm 1.

We are given our target distribution is the posterior density of μ, ϕ of a normal under standard reference priors $p(\mu, \phi) \propto 1/\phi$, hence

$$f(\mu, \phi) = p(\mu, \phi \mid X) \propto \phi^{\frac{n}{2}-1} \exp\left\{ -\frac{\phi}{2} \sum_{i=1}^{n} (x_i - \mu)^2 \right\}$$

With two parameters, we employ two proposal functions that fit the constraints of each parameter, both symmetric, such that our acceptance ratio will only depend on the ratio of the target distribution.

For the proposal distribution of μ, we propose a symmetric normal proposal under some stepsize hyperparameter ψ_μ, which we tune later. For the proposal distribution of ϕ, we take the $\ln \phi$, and we propose a random walk on $\ln \phi$ under a normal, with stepsize hyperparameter ψ_ϕ, also subject to hyperparmeter tuning. We take the log as the variance and precision parameter must always be normal. Hence, to employ a symmetric proposal, we must do it in log-space to ensure positivity of the generated proposal values. Hence, this proposal will be symmetric in log-space, which we evaluate the acceptance probability in. Hence, we have the following set of proposals.

$$\begin{aligned} g(\mu^* \mid \mu) &= \frac{1}{\sqrt{2\pi}\psi_\mu} \exp\left(-\frac{1}{2\psi_\phi^2} (\mu^* - \mu)^2 \right) \\ \ln(\phi^*) \mid \ln(\phi) &= \ln \phi + \epsilon, \quad \epsilon \sim N(0, \psi_\phi^2) \\ \implies g(\ln(\phi^*) \mid \ln(\phi)) &= \frac{1}{\sqrt{2\pi}\psi_\phi} \exp\left(-\frac{1}{2\psi_\phi^2} (\ln(\phi^*) - \ln(\phi))^2 \right) \end{aligned}$$

We evaluate the acceptance ratio as follows,

$$\frac{f(\boldsymbol{\theta}^*)}{f(\boldsymbol{\theta}_t)} \frac{g(\boldsymbol{\theta}_t \mid \boldsymbol{\theta}^*)}{g(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}_t)}$$

where we let $\boldsymbol{\theta} = (\mu, \phi)$ and $g(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}) = g(\mu^* \mid \mu) g(\ln(\phi^*) \mid \phi)$.

Since both proposals are normal, hence symmetric, we realize our acceptance ratio is simply the ratio of the posteriors.

$$\frac{f(\boldsymbol{\theta}^*)}{f(\boldsymbol{\theta}_t)} \frac{g(\boldsymbol{\theta}_t \mid \boldsymbol{\theta}^*)}{g(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}_t)} \propto \frac{f(\boldsymbol{\theta}^*)}{f(\boldsymbol{\theta}_t)} \frac{\exp\left(-\frac{1}{2\psi_\phi^2}(\mu - \mu^*)^2\right) \exp\left(-\frac{1}{2\psi_\phi^2}(\ln(\phi) - \ln(\phi^*))^2\right)}{\exp\left(-\frac{1}{2\psi_\phi^2}(\mu^* - \mu)^2\right) \exp\left(-\frac{1}{2\psi_\phi^2}(\ln(\phi^*) - \ln(\phi))^2\right)} = \frac{f(\boldsymbol{\theta}^*)}{f(\boldsymbol{\theta}_t)}$$

Hence, we present the pseudocode below.

Algorithm 1 Metropolis algorithm to approximate $\mu, \phi \mid X$

Require: $f(x) \propto \phi^{\frac{n}{2}-1} \exp\left(-\frac{\phi}{2}\sum_{i=1}^n (x_i - \mu)^2\right)$ (target distribution), $g(\mu^*|\mu) \propto \exp\left(-\frac{1}{2\psi_\phi^2}(\mu^* - \mu)^2\right)$, $g(\ln\phi^*|\ln\phi) \propto \exp\left(-\frac{1}{2\psi_\phi^2}(\ln(\phi^*) - \ln(\phi))^2\right)$ (proposal distributions), T (number of iterations)

Initialize μ_0 and ϕ_0
for $t = 0$ to T **do**
    Draw μ^*, ϕ^* from $g(\mu^*|\mu_t)$, $g(\ln\phi^*|\ln\phi_t)$
    Compute acceptance ratio:

$$r = \frac{f(\mu^*, \phi^*)}{f(\mu_t, \phi_t)} \frac{g(\mu_t|\mu^*) g(\ln\phi_t|\ln\phi^*)}{g(\mu^*|\mu_t) g(\ln\phi^*|\ln\phi)} = \frac{f(\mu^*, \phi^*)}{f(\mu_t, \phi_t)}$$

    Compute acceptance probability $\alpha = \min(1, r)$
    Draw $u \sim \text{uniform}(0, 1)$
    if $u < \alpha$ **then**
        $\mu_{t+1} \leftarrow \mu^*$, $\phi_{t+1} \leftarrow \phi^*$
    else if $u \geq \alpha$ **then**
        $\mu_{t+1} \leftarrow \mu_t$, $\phi_{t+1} \leftarrow \phi_t$
    end if
end for

Part 2

We implement two versions of the Metropolis sampler, the 'vanilla' Metropolis-Hastings sampler, and the Gibbs sampler, which we have the pseudocode in Algorithm 2. In the Gibbs sample, we now take each proposal for each parameter separately, conditional on the previous accepted proposal value. That is, the Gibbs sample is in a more specific sense, sampling each parameter conditional on the accepted proposal value of all other parameter values.

As mentioned, we implement the code for normal data generated from $N(200, 0.5)$, under different sample sizes of $N = 10, 30, 100$. We initialize the samplers at $\mu_0 = 0$ and $\phi_0 = 5$.

The codes are available online, where we note due to overflow constraints, the acceptance ratio is evaluated in log-scale `https://github.com/jeonghlee12/STAT5114/tree/main/MCMC`

Algorithm 2 Gibbs sampler algorithm to approximate $\mu, \phi \,|\, X$

Require: Same as Algorithm 1

Initialize μ_0 and ϕ_0
for $t = 0$ to T **do**
  Draw μ^*, ϕ^* from $g(\mu^*|\mu_t)$
  Compute acceptance ratio:

$$r_\mu = \frac{f(\mu^*, \phi_t)}{f(\mu_t, \phi_t)} \frac{g(\mu_t|\mu^*)}{g(\mu^*|\mu_t)} = \frac{f(\mu^*, \phi_t)}{f(\mu_t, \phi_t)}$$

  Compute acceptance probability $\alpha_\mu = \min(1, r_\mu)$
  Draw $u \sim \text{uniform}(0, 1)$
  if $u < \alpha_\mu$ **then**
    $\mu_{t+1} \leftarrow \mu^*$
  else if $u \geq \alpha$ **then**
    $\mu_{t+1} \leftarrow \mu_t$
  end if
  Draw $g(\ln \phi^* | \ln \phi_t)$
  Compute acceptance ratio:

$$r_\phi = \frac{f(\mu_{t+1}, \phi^*)}{f(\mu_{t+1}, \phi_t)} \frac{g(\ln \phi_t | \ln \phi^*)}{g(\ln \phi^* | \ln \phi)} = \frac{f(\mu_{t+1}, \phi^*)}{f(\mu_{t+1}, \phi_t)}$$

  Compute acceptance probability $\alpha_\phi = \min(1, r_\phi)$
  Draw $u \sim \text{uniform}(0, 1)$
  if $u < \alpha_\phi$ **then**
    $\phi_{t+1} \leftarrow \phi^*$
  else if $u \geq \alpha$ **then**
    $\phi_{t+1} \leftarrow \phi_t$
  end if
end for

Part 3

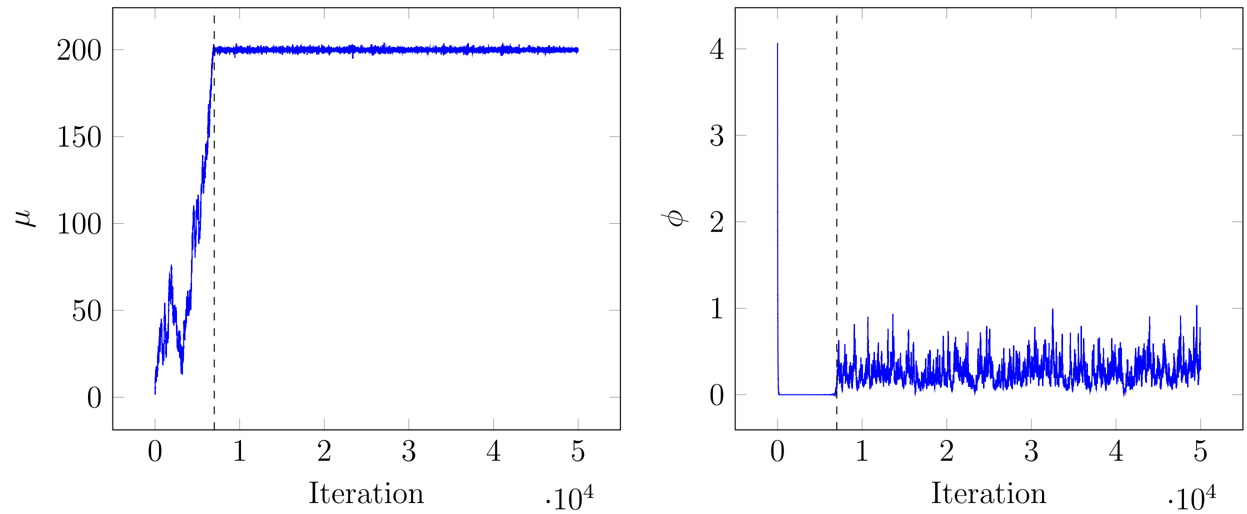
Here, we present summaries of the simulations under $N = 10, 30, 100$, and under both samplers. We provide the trace plots, with which we identify the burn-in period, histograms of the marginal posteriors after burn-in, and the contour plots of the generated samples after burn-in.

We generate a total of 50000 samples to accommodate for potential long burn-in times. For the Metropolis-Hastings sampler, we used hyperparameters $\psi_\mu = 1$ and $\psi_\phi = 0.1$, and for the Gibbs sampler, we used hyperparameters $\psi_\mu = 0.5$ and $\psi_\phi = 0.1$.

$N = 10$

Metropolis-Hastings

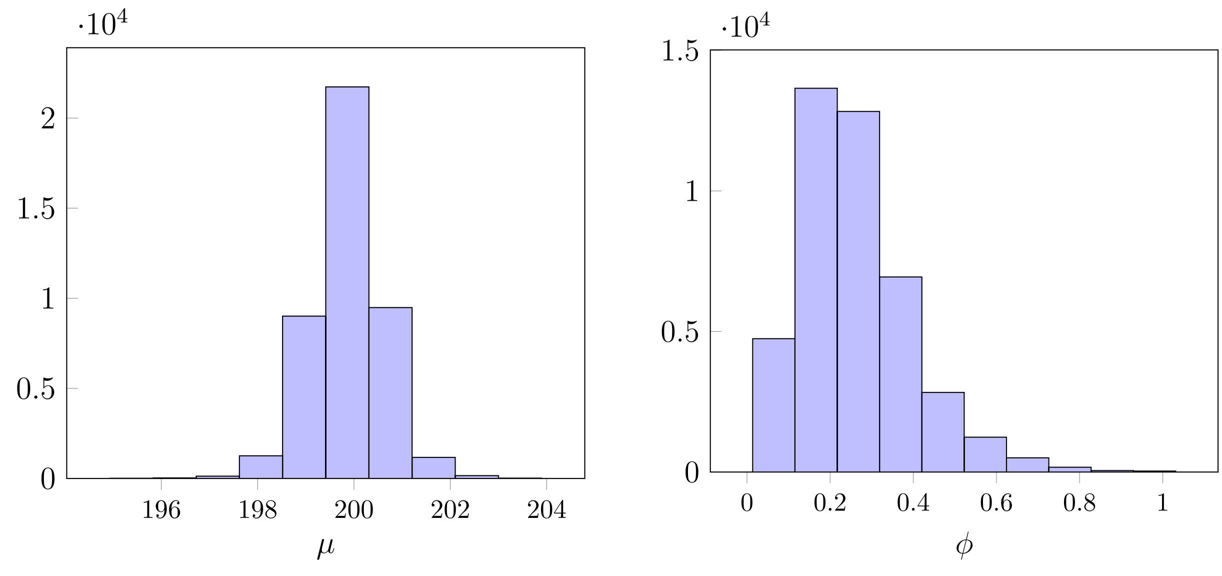
We identify under the trace plots in Figure 1, the burn-in period for the Metropolis-Hastings sampler under $N = 10$ to be about 7000 iterations. Total acceptance rate was about 62%, a reasonable value.



(a) Trace plot for μ (b) Trace plot for ϕ

Figure 1: Trace plots for Metropolis-Hastings $N = 10$ (Accept. rate: 0.62)

Figure 2 shows the histogram of the marginal posteriors of each parameter after the burn-in.



(a) Histogram for μ (b) Histogram for ϕ

Figure 2: Histograms for Metropolis-Hastings $N = 10$

Lastly, we display the contour plots of the joint samples $\mu, \phi \mid X$ after burn-in in Figure 3

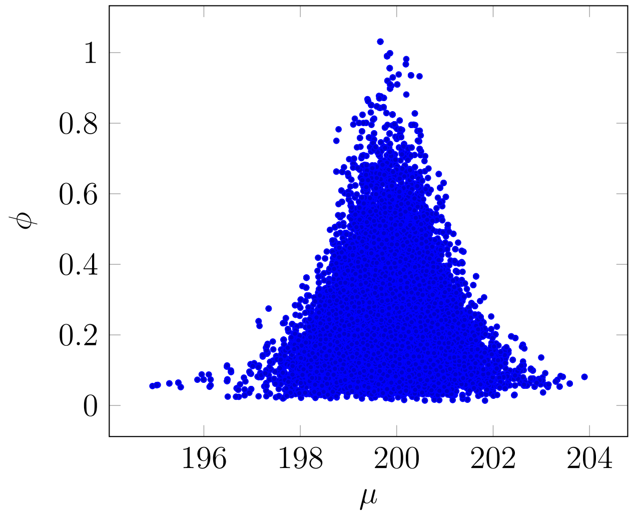
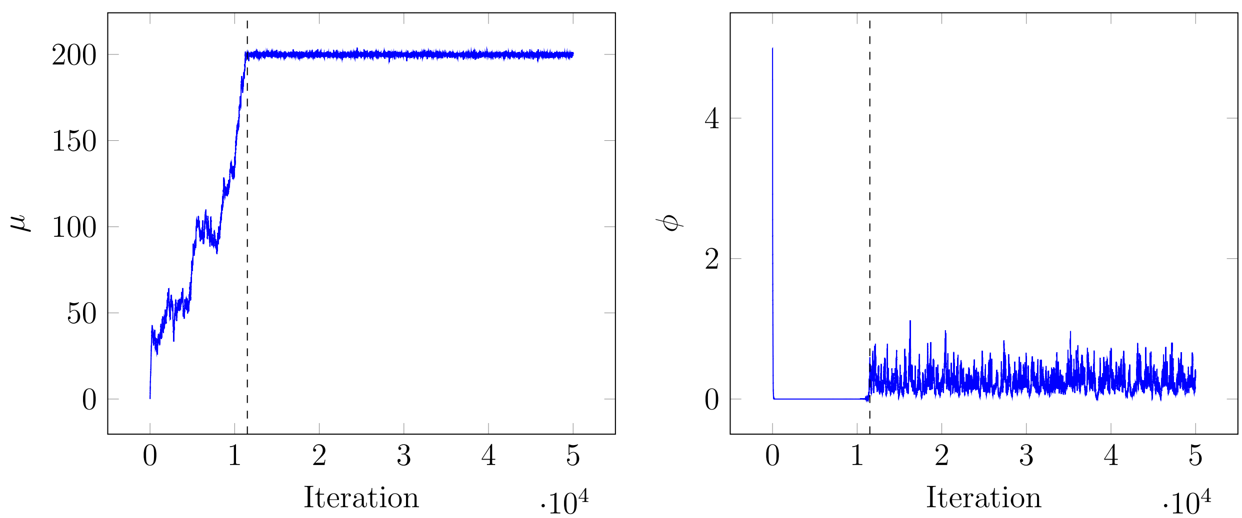


Figure 3: Contour plot of the joint sample for Metropolis-Hastings $N = 10$

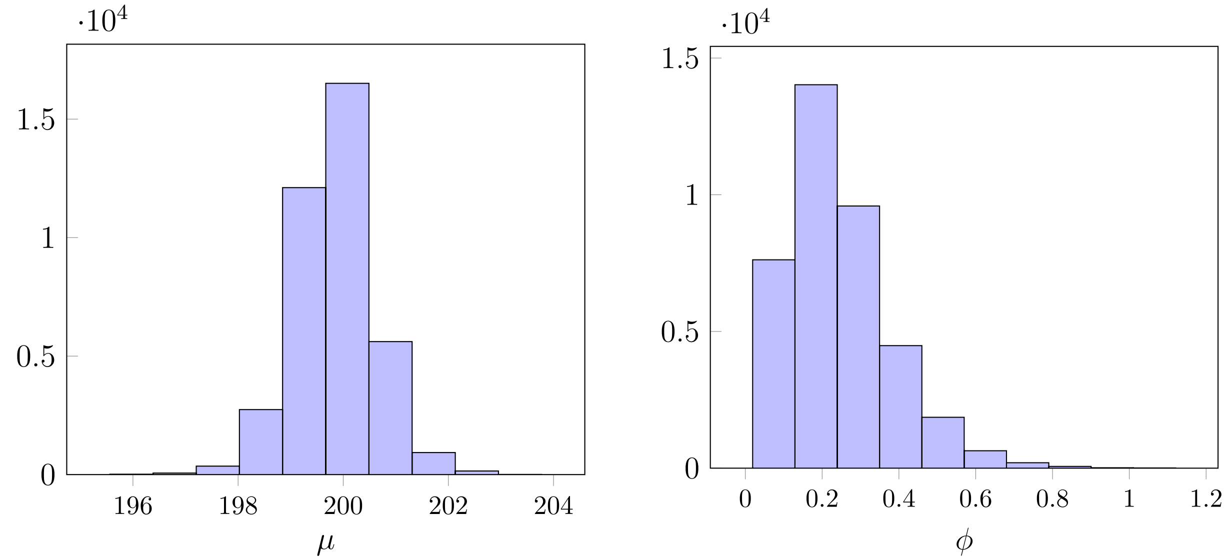
Gibbs Sampling

We now provide similar plots for the generated sample under Gibb's method, with a burn-in time of 11500 iterations identified. The acceptance rate for μ was about 82% and for ϕ was 93%.



(a) Trace plot for μ (Accept. rate: 0.82) (b) Trace plot for ϕ (Accept. rate: 0.93)

Figure 4: Trace plots for Gibbs sampling $N = 10$



(a) Histogram for μ

(b) Histogram for ϕ

Figure 5: Histograms for Gibbs sampling $N = 10$

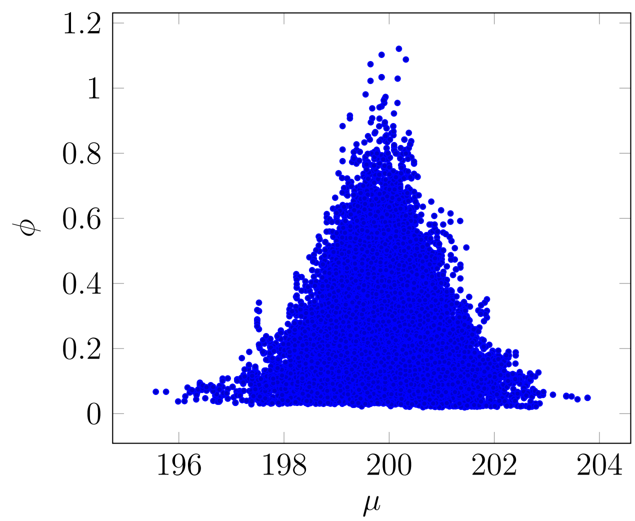
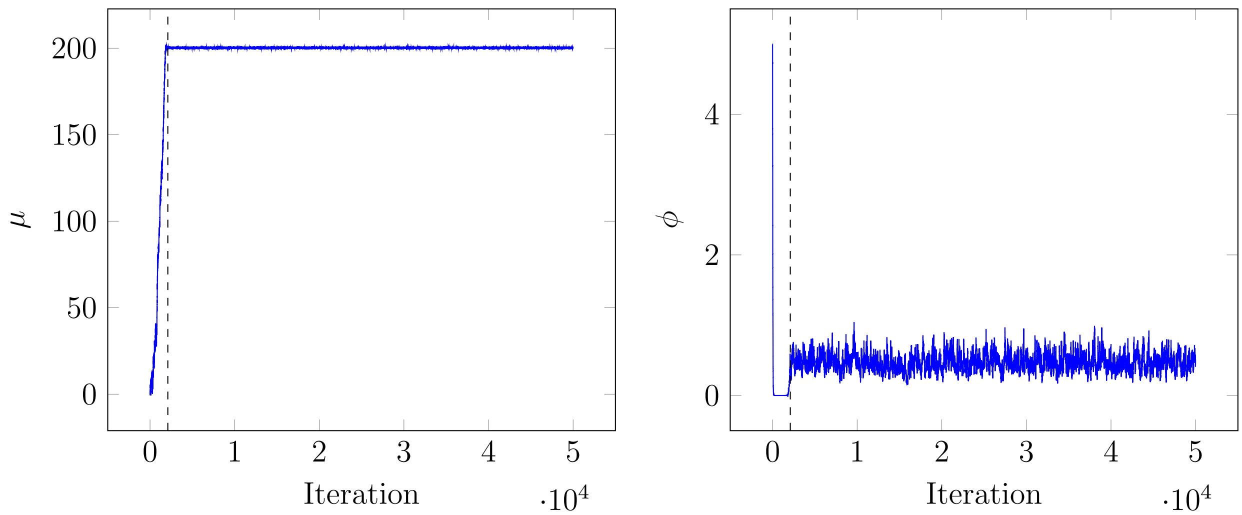


Figure 6: Contour plot of the joint sample for Gibbs sampling $N = 10$

$N = 30$

Metropolis-Hastings

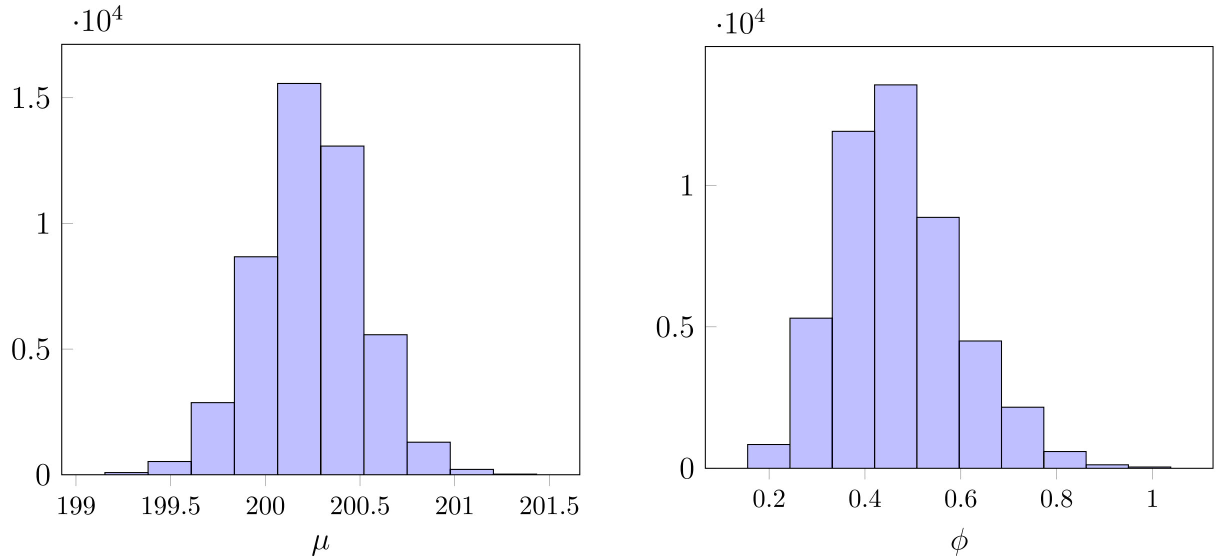
We identify under the trace plots in Figure 7, the burn-in period for the Metropolis-Hastings sampler under $N = 30$ to be about 2100 iterations. Total acceptance rate was about 32%.



(a) Trace plot for μ (b) Trace plot for ϕ

Figure 7: Trace plots for Metropolis-Hastings $N = 30$ (Accept. rate: 0.32)

Figure 8 displays the marginal posterior distributions.



(a) Histogram for μ (b) Histogram for ϕ

Figure 8: Histograms for Metropolis-Hastings $N = 30$

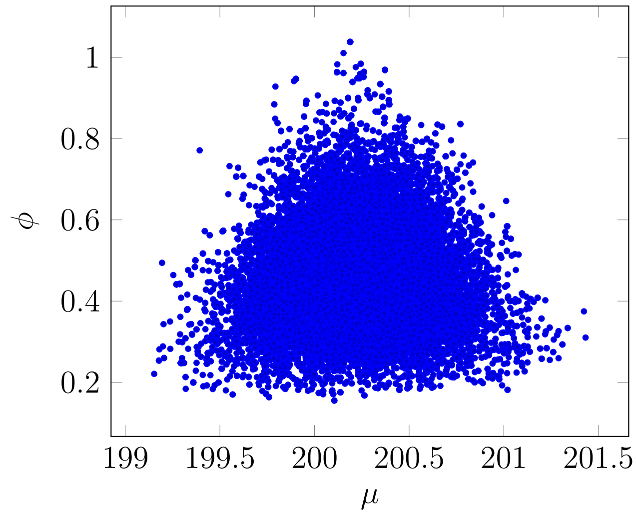
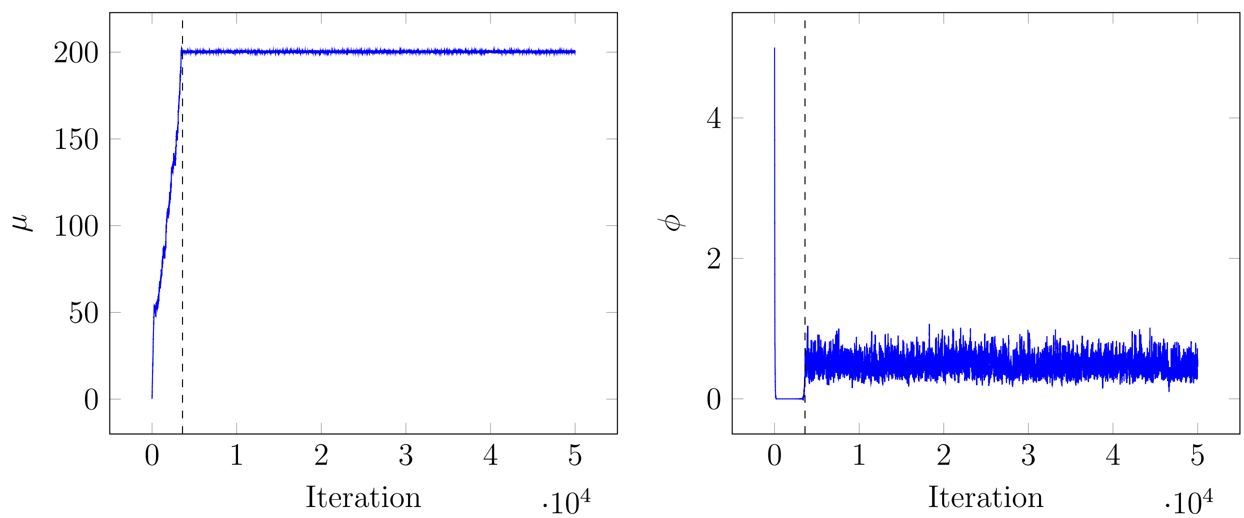


Figure 9: Contour plot of the joint sample for Metropolis-Hastings $N = 30$

Gibbs Sampling

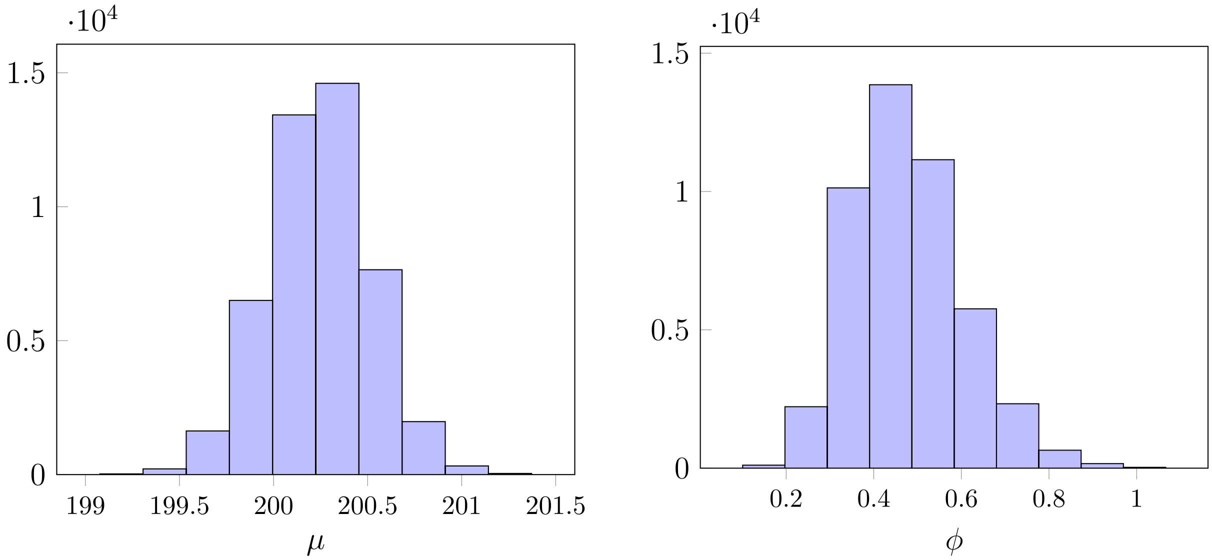
We now provide similar plots for the generated sample under Gibb's method, with a burn-in time of 3600 iterations identified. The acceptance rate for μ was about 55% and for ϕ was 88%.



(a) Trace plot for μ (Accept. rate: 0.55)

(b) Trace plot for ϕ (Accept. rate: 0.88)

Figure 10: Trace plots for Gibbs sampling $N = 30$



(a) Histogram for μ

(b) Histogram for ϕ

Figure 11: Histograms for Gibbs sampling $N = 30$

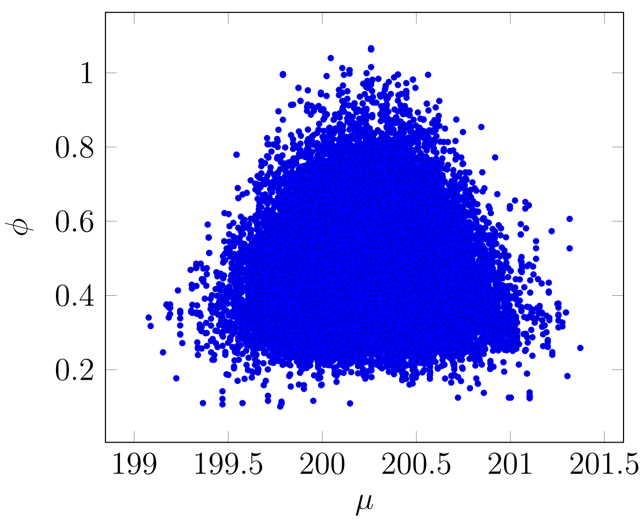
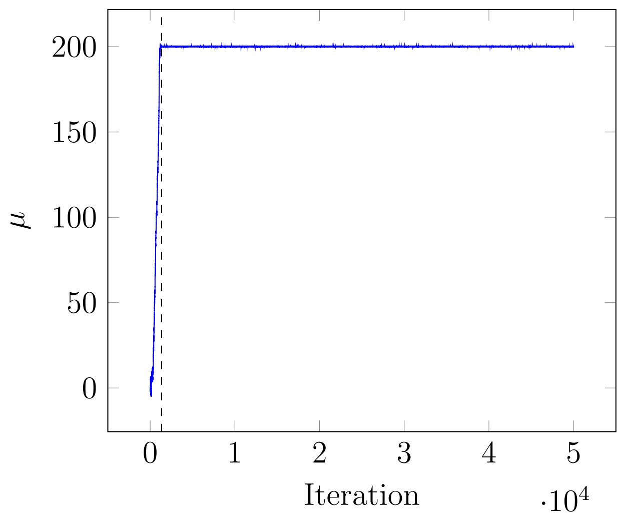


Figure 12: Contour plot of the joint sample for Gibbs sampling $N = 30$

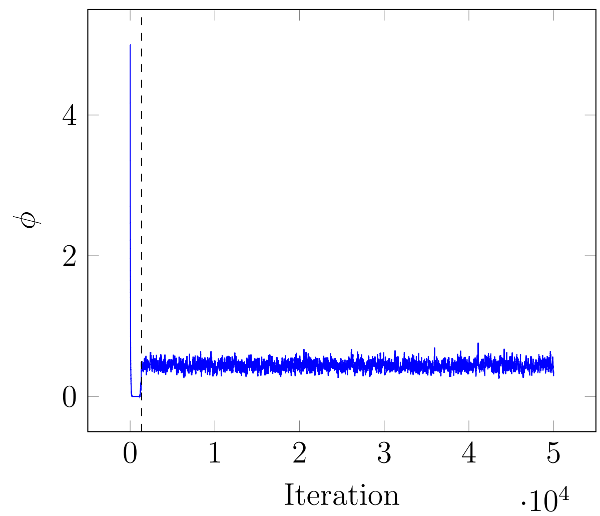
$N = 100$

Metropolis-Hastings

Lastly, we identify under the trace plots in Figure 13, the burn-in period for the Metropolis-Hastings sampler under $N = 100$ to be about 1350 iterations. Total acceptance rate was about 18%.



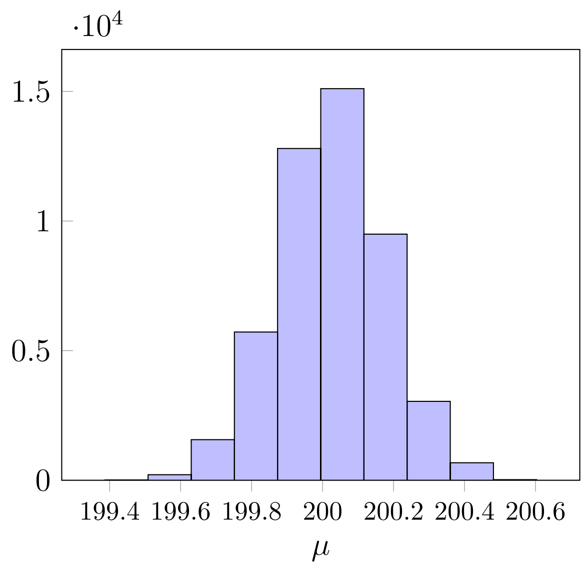
(a) Trace plot for μ



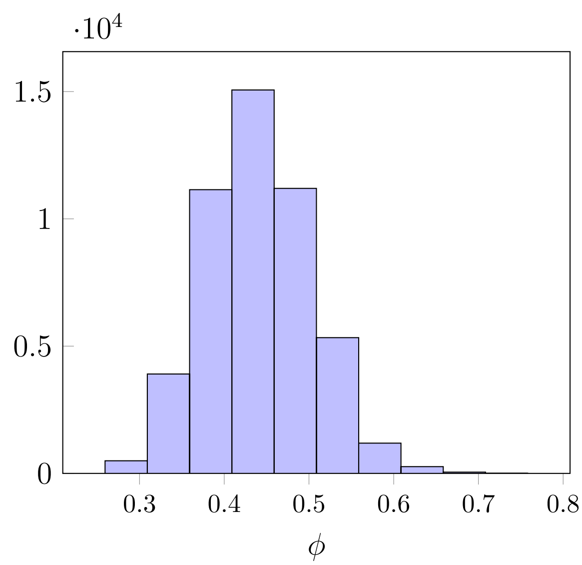
(b) Trace plot for ϕ

Figure 13: Trace plots for Metropolis-Hastings $N = 100$ (Accept. rate: 0.18)

Figure 14 displays the marginal posterior distributions.



(a) Histogram for μ



(b) Histogram for ϕ

Figure 14: Histograms for Metropolis-Hastings $N = 100$

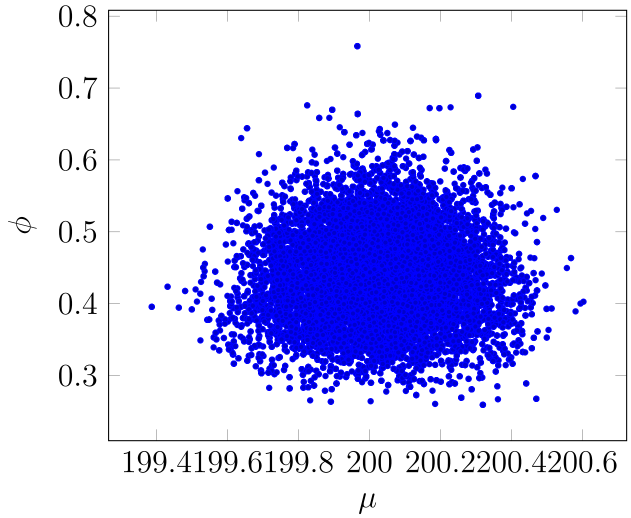
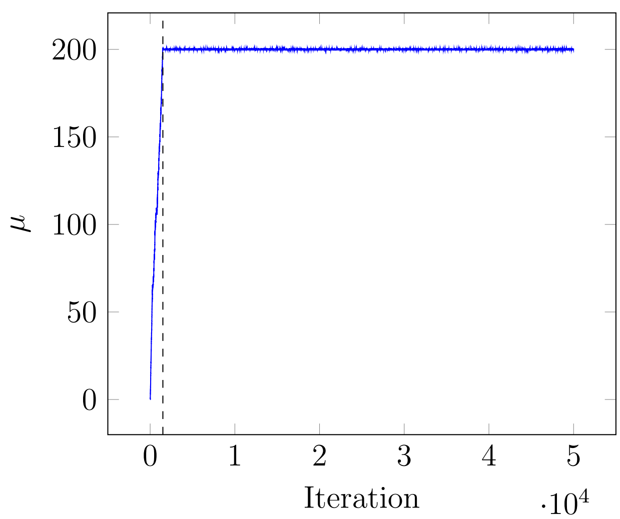


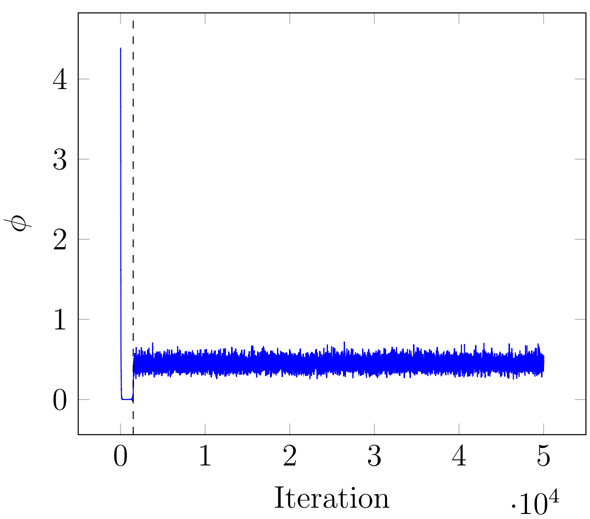
Figure 15: Contour plot of the joint sample for Metropolis-Hastings $N = 100$

Gibbs Sampling

We now provide similar plots for the generated sample under Gibb's method, with a burn-in time of 1500 iterations identified. The acceptance rate for μ was about 36% and for ϕ was 79%.

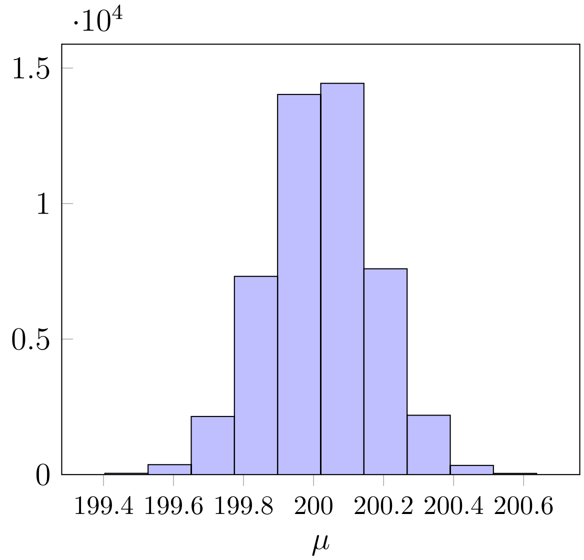


(a) Trace plot for μ (Accept. rate: 0.36)

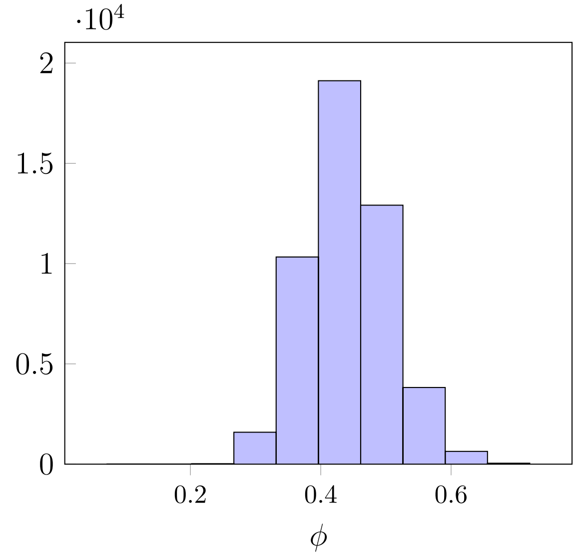


(b) Trace plot for ϕ (Accept. rate: 0.79)

Figure 16: Trace plots for Gibbs sampling $N = 100$



(a) Histogram for μ



(b) Histogram for ϕ

Figure 17: Histograms for Gibbs sampling $N = 100$

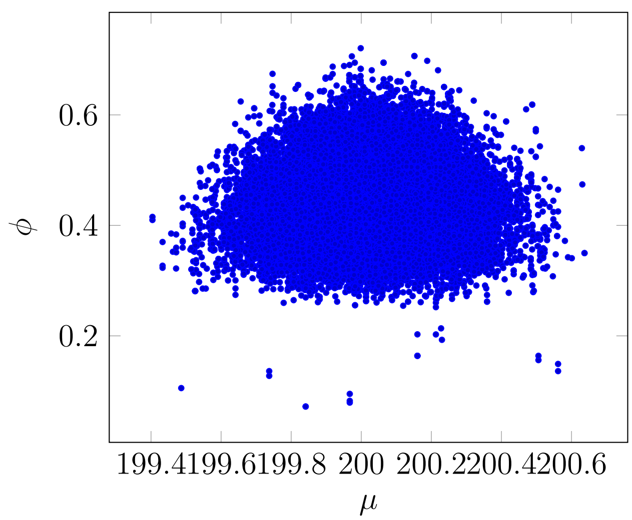


Figure 18: Contour plot of the joint sample for Gibbs sampling $N = 100$