

2019-11-27

computational Statistics

HW#9

192STG11 우나영

computational Statistics

HW#9

1. Problem

Markov Chain and MCMC

유한(finite) 혹은 가산 무한한(countably infinite) 값을 가지는 확률 변수 $X^{(t)}$ 의 수열 $\{X^{(t)}\}, t = 0, 1, \dots$ 을 생각해보자. 이때 $X^{(t)}$ 의 값은 state 이며 모든 가능한 state 들의 집합을 S 이다. $X^{(0)}, \dots, X^{(t)}$ 의 joint distribution 으로 표현하면 $P[X^{(0)}, \dots, X^{(t)}] = P[X^{(t)} | x^{(0)}, \dots, x^{(t-1)}] \times P[X^{(t-1)} | x^{(0)}, \dots, x^{(t-2)}] \times \dots \times P[X^{(1)} | x^{(0)}] \times P[X^{(0)}]$ 이다. 이때 Markov property 는 각 시점의 state 가 이전의 state 에만 dependent 한 것을 말한다. 이러한 Markov property 를 만족하는 수열을 Markov chain 이라 말하며 $P[X^{(0)}, \dots, X^{(t)}] = P[X^{(t)} | x^{(t-1)}] \times P[X^{(t-1)} | x^{(t-2)}] \times \dots \times P[X^{(1)} | x^{(0)}] \times P[X^{(0)}]$ 로 표현할 수 있다. $X^{(t)}$ 의 state 가 시점에 t 일 때 i 에서 시점 $t+1$ 일 때 j 로 변화할 확률은 전이확률(transition probability)이며 $p_{ij}^{(t)} = P[X^{(t+1)} = j | X^{(t)} = i], t = 0, 1, \dots$ and $x^{(0)}, x^{(1)}, \dots, x^{(t-1)}, i, j \in S$ 로 정의된다.

Markov Chain 이 unique 한 stationary distribution 가지기 위해서는 irreducible 하고 aperiodic 해야 한다. Irreducible 이란 어떤 state j 에서든 유한한 시점 안에 다른 모든 state 로 이동이 가능한 것을 말한다. 즉, $\text{for all } i, j \in S, P[X^{(t+n)} = j | X^{(t)} = i] > 0$ 이다. Aperiodic 은 어떤 상태 i 에서 출발하여, 다시 상태 i 로 돌아올 때까지의 주기가 없는 경우를 말한다. 이러한 irreducible 하고 aperiodic 하며 stationary distribution π 를 갖는 Markov chain 의 실현치들은 $X^{(n)}$ 이 in distribution 으로 π 에 수렴하며 함수 h 에 대해 다음과 같은 성질을 만족한다. 이는 일반화된 강대수 법칙(SLLN)인 ergodic theorem 의 형태 중 하나이다. 즉, ergodic 은 SLLN 과 달리 $X^{(t)}$ 가 iid 라는 가정이 없어도 expected of $h(x)$ 로 수렴하는 성질을 말한다.

$$\frac{1}{n} \sum_{t=1}^n h(X^{(t)}) \rightarrow E_{\pi}\{h(X)\} \text{ a.s. as } n \rightarrow \infty \text{ when provided } E_{\pi}\{|h(X)|\}$$

MCMC sampling 은 stationary distribution 이 target distribution f 로 같아지도록 irreducible 하고 aperiodic 한 Markov chain 을 만드는 방법이다. MCMC에서 Markov chain 을 만드는 두가지 방법으로 Metropolis-Hastings 와 Gibbs sampling 을 소개하겠다.

Metropolis-Hastings

첫 번째로 Metropolis-Hastings 는 $t=0$ 일 때 proposal density g 를 따르는 $X^{(0)} = x^{(0)}$ 에서 시작한다. 이때 target density 를 f 라 할 때 $f(x^{(0)}) > 0$ 이다. $X^{(t)} = x^{(t)}$ 가 주어졌을 때 $X^{(t+1)}$ 생성은 아래와 같은 과정을 거친다.

Step1) 후보 X^* 를 proposal density $g(\cdot | x^{(t)})$ 을 생성한다.

Step2) Metropolis-Hastings Ratio $MHR = \frac{f(X^*)g(x^{(t)}|X^*)}{f(x^{(t)})g(X^*|x^{(t)})}$

Step3) $X^{(t+1)} = \begin{cases} X^* & w.p \min\{R(x^{(t)}, X^*), 1\} \\ x^{(t)} & o.w. \end{cases}$

Step4) t 를 증가시키고 step1 으로 돌아간다.

Independence Chain

Metropolis-Hastings 알고리즘의 Proposal density g 를 $g(X^*|X^{(t)}) = g(X^*)$ 로 가정하면 Independence MC(Markov Chain)을 생성할 수 있다. 이러한 가정하에 Target density 가 f 일 때 Metropolis-Hastings Ratio(MHR)는 아래와 같다.

$$MH \text{ Ratio} : R(X^{(t)}, X^*) = \frac{f(X^*)g(X^{(t)})}{f(X^{(t)})g(X^*)}$$

$f(x) > 0$ 일 때 $g(x)$ 가 0 이라면 위의 결과로 생성된 MC 는 irreducible 하고 aperiodic 한 성질을 만족한다. 또한 MHR 은 Importance ratio 의 ratio 로 표현할 수 있다. $w^* = f(X^*)/g(X^*)$ 이고 $w^{(t)} = f(X^{(t)})/g(X^{(t)})$ 이면 MH ratio $R(X^{(t)}, X^*) = \frac{w^*}{w^{(t)}}$ 로 표현된다. 만약 $w^{(t)}$ 가 w^* 보다 크면 MHR 은 작아질 것이고 update 가 발생하지 않아 chain 은 현재의 값에 고착될 것이다. 이를 방지하기 위해서 X^* 를 생성하는 proposal density g 가 target density 인 f 와 비슷한 동시에 f 보다 양 끝단 tail 이 heavier 해야 한다.

Bayesian Inference

Metropolis-Hastings 와 같은 MCMC 알고리즘은 Bayesian 추론의 훌륭한 도구이다. Bayesian 추론의 핵심은 관측치 y 를 바탕으로 가정한 통계 모형 likelihood 함수 $L(\theta|y)$ 와 parameter θ 의 prior 인 $p(\theta)$ 에 비례하는 θ 의 posterior $p(\theta|y) = c \times p(\theta) \times L(\theta|y)$ 를 추정하는 것이다. 이때 c 는 $p(\theta|y)$ 의 적분 값을 1 으로 만들어주는 normalizing constant 로 θ 에 대한 직접적인 추론을 하기 위해서는 c 를 구해야

한다. 그러나 c 는 수학적으로 계산하기 매우 까다롭다. 따라서 c 를 계산하는 대신에 MCMC 알고리즘을 이용해 stationary density 가 posterior density of θ 인 MC 을 구한다면 샘플링을 이용해 쉽게 θ 에 대한 통계적 추론을 할 수 있다.

Stationary density 가 Posterior density 인 Independence chain 을 생성하기 위한 Metropolis-Hastings set up 은 아래와 같다.

$$f(\theta) = p(\theta|y) \text{ posterior as target density}$$

$$g(\theta^*) = p(\theta^*) \text{ prior as proposal density}$$

$$\Rightarrow R(\theta^{(t)}, \theta^*) = \frac{p(\theta^*|y)p(\theta^{(t)})}{p(\theta^{(t)}|y)p(\theta^*)} = \frac{cp(\theta^*)L(\theta^*|y)p(\theta^{(t)})}{cp(\theta^{(t)})L(\theta^{(t)}|y)p(\theta^*)} = \frac{L(\theta^*|y)}{L(\theta^{(t)}|y)} : \text{Likelihood ratio}$$

즉, 현재인 $\theta^{(t)}$ 의 likelihood 와 proposal 인 θ^* 의 likelihood 비율의 확률로 θ^* 로 update 한다.

Random Walk Chains

랜덤 워크 체인은 Metropolis-Hastings 를 응용해 Markov Chain 을 만드는 방법이다. ϵ 을 h 라는 density 에서 뽑아 $X^* = x^{(t)} + \epsilon$ 를 생성한다. 이때 h 는 일반적으로 표준 정규 분포 또는 표준화 t 분포를 사용한다. 이러한 경우 proposal density 인 $g(x^*|x^{(t)}) = h(x^* - x^{(t)})$ 이다. 만약 target 함수인 f 의 범위가 연결되었으며 h 가 o 근방에서 o 이상이라면 이러한 방법으로 생성한 chain 은 *irreducible* 하고 *aperiodic* 하다.

Gibbs Sampling

두 번째로 MCMC 를 만드는 Gibbs Sampling 은 단변량의 조건부 확률분포(univariate conditional density)에서 순차적으로 샘플을 생성하는 방법이다. $X = (X_1, \dots, X_p)^T$ 이고 $X_{-i} = (X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_p)^T$ 라 할 때, $X_i|X_{-i} = x_{-i}$ 의 단변량 조건부 확률분포인 $f(x_i|x_{-i})$ for $i = 1, \dots, p$ 는 쉽게 샘플을 생성할 수 있다.

Step1. $t=0$ 에서 초기치로 $x^{(0)}$ 을 설정한다.

Step2. 차례로 샘플을 생성시킨다.

$$X_1^{(t+1)} | \cdot \sim f(x_1|x_2^{(t)}, \dots, x_p^{(t)})$$

$$X_2^{(t+1)} | \cdot \sim f(x_2|x_1^{(t+1)}, x_3^{(t)}, \dots, x_p^{(t)})$$

$$\vdots$$

$$X_{p-1}^{(t+1)} | \cdot \sim f(x_{p-1} | x_1^{(t+1)}, x_2^{(t+1)}, \dots, x_{p-2}^{(t+1)}, x_p^{(t)})$$

$$X_p^{(t+1)} | \cdot \sim f(x_p | x_1^{(t+1)}, x_2^{(t+1)}, \dots, x_{p-1}^{(t+1)})$$

Step3. t 를 증가시키고 step2 로 돌아간다.

2. Result

HW1. Example 7.2 (Mixture Distribution) using Independence Chain

$$y_1, \dots, y_{100} \sim iid \delta N(7, 0.5^2) + (1 - \delta) N(10, 0.5^2)$$

y_1, \dots, y_{100} 은 iid 로 GMM(Gaussian Mixture Model)을 따른다. δ 를 추정하기 위해 Metropolis-Hastings 알고리즘으로 stationary density 가 posterior density of δ 가 되는 MCMC chain 을 생성할 것이다. 관측 데이터 $y_i, i = 1, \dots, 100$ 는 $\delta = 0.7$ 인 GMM 에서 생성된 것이므로 posterior density of δ 가 0.7 에 집중되었는지 확인해보아야 한다. δ 의 prior 가 Beta(1,1) 그리고 Beta(2,10)인 두 개의 independence chain 을 생성하여 각각의 chain 의 stationary density 를 비교해 볼 것이다. initial value 인 $\delta[1]$ 는 Unif(0,1)에서 생성한 값이다.

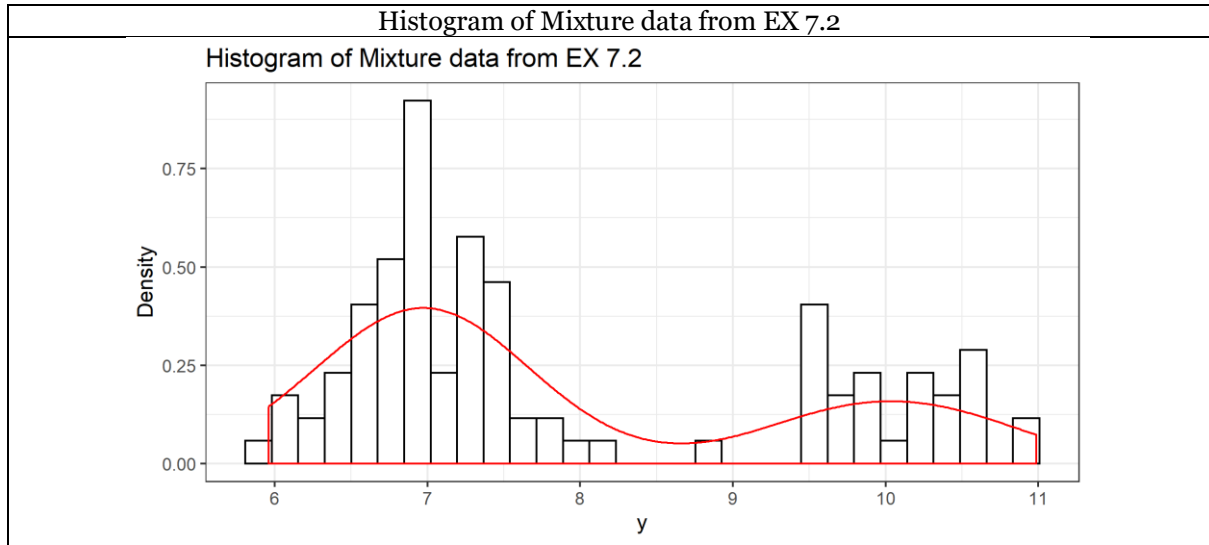
$$\text{Independence chain } \begin{cases} 1) \text{ with prior Beta}(1,1) \\ 2) \text{ with prior Beta}(2,10) \end{cases}$$

Step1) 후보 δ^* 는 δ 의 prior 인 Beta(1,1) (Beta(2,10))에서 생성한다.

$$\text{Step2) Metropolis Hastings } R = \frac{L(\delta^*|y)}{L(\delta^{(t)}|y)}$$

$$\text{Step3) } \delta^{(t+1)} = \begin{cases} \text{get } \delta^* \text{ w.p } R & \text{if } \min(R, 1) \\ \text{get } \delta^{(t)} & \text{o.w} \end{cases}$$

Step4) 위의 과정을 샘플의 개수인 n 번 반복한다.



Proposal density Sample Path	Beta(1,1)	Beta(2,10)
	<p>Sample Path for delta from Beta(1,1)</p>	<p>Sample Path for delta from Beta(2,10)</p>
Histogram of $\delta^{(t)}$ after burn in the first 200	<p>Histogram of delta generated from Beta(1,1)</p>	<p>Histogram of delta generated from Beta(2,10)</p>
min	0.5395	0.6241
Q1	0.6661	0.6241
Q2	0.6953	0.6280
mean	0.6964	0.6366
Q3	0.7254	0.6304
max	0.8287	0.7022

Sample Path plot 은 iteration number(t)에 따른 δ 의 양상을 보여준다. 이를 통해 Markov Chain 이 starting value 를 벗어나 δ 의 posterior 의 sample 을 생성하는지 알 수 있다. Beta(1,1)을 proposal로 사용했을 경우 starting value를 빠르게 벗어나며 MC가 stationary로 수렴하는 양상을 보이는 반면 Beta(2,10)가 proposal 인 경우 starting value 를 느리게 벗어나며 update 가 거의 발생하지 않아 계단식 양상을 보여준다. 즉, MC가 nonstationary 하므로 Beta(2,10)은 proposal 로 적합하지 않다는 것을 알 수 있다. 각각의 proposal 에서 생성된 샘플의 초반 200 개를 burn-in 한 히스토그램을 비교해 보았을

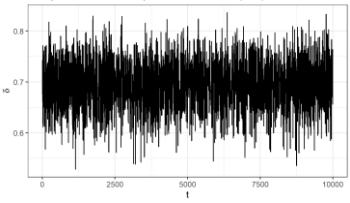
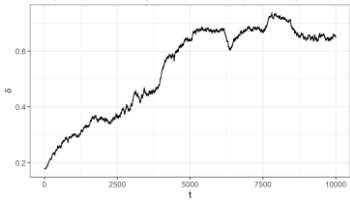
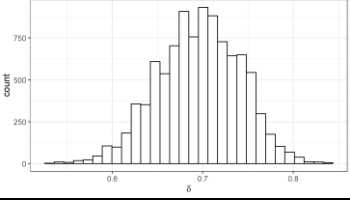
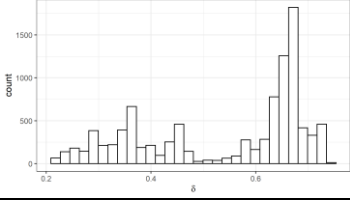
때 proposal 이 Beta(1,1)일 경우 delta 의 mean 값이 delta 의 true 값인 0.7 에 수렴하지만 proposal 이 Beta(2,10)인 경우 delta 의 분포가 산발적이기 때문에 신뢰성 있는 chain 이라 할 수 없다.

HW2. Example 7.2 (Mixture Distribution) using Random Walk Chain

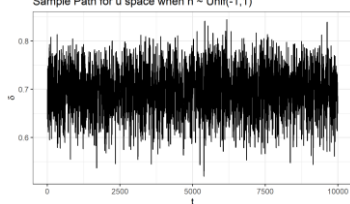
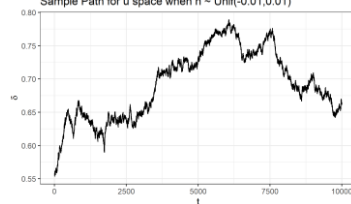
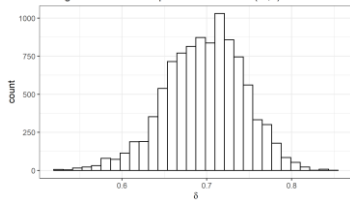
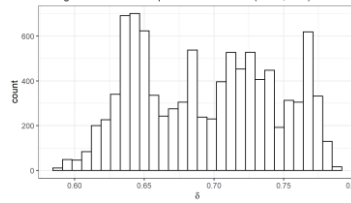
Prior 가 Unif(0,1)인 δ 의 posterior 의 random walk chain 을 형성할 때, current $\delta^{(t)}$ 에 Unif(-a, a)에서 뽑은 random number 를 증가시켜 proposal 을 생성한다. Chain 이 만들어지는 과정 중에 proposal $\delta^* \notin [0,1]$ 인 경우 $\delta^{(t+1)}$ 이 δ^* 으로 step 을 옮기지 못하게 막거나 reparameterizing 하는 방법이 있다.

$U = \text{logit}(\delta) = \log \frac{\delta}{1-\delta}$ 라 하자. Reparameterizing 의 두 가지 관점이 있다. 첫 번째는 δ space 에서 chain 을 만드는 것이다. δ^* 의 Metropolis-Hastings ratio 는 $\frac{f(\delta^*)g(\text{logit}\{\delta^{(t)}\})|\text{logit}\{\delta^*\}|J(\delta^{(t)})}{f(\delta^{(t)})g(\text{logit}\{\delta^*\})|\text{logit}\{\delta^{(t)}\}|J(\delta^*)}$ 이다. 이때, $|J(\delta^{(t)})|$ 는 δ 에서 u 로 변환할 때 발생하는 Jacobian term 이다. 두 번째는 u space 에서 chain 을 만드는 것이다. 이 경우에 δ 의 target density 가 $\delta = \text{logit}^{-1}\{U\} = \exp\{U\} / (1 + \exp\{U\})$ 변환된다. $U^* = u^*$ 일 때, Metropolis-Hastings ratio 는 $\frac{f(\text{logit}^{-1}\{u^*\})|J(u^*)|g(u^{(t)}|u^*)}{f(\text{logit}^{-1}\{u^{(t)}\})|J(u^{(t)})|g(u^*|u^{(t)})}$ 이다. $|J(u^*)| = 1/|J(\delta^*)|$ 이기 때문에 위의 두 가지 방법으로 만든 random chain 은 동일하다.

Method 1) δ - space

Proposal density	Unif(-1,1)	Unif(-0.01,0.01)
Sample Path		
Histogram of $\delta^{(t)}$ after burn in the first 200		
min	0.5286	0.2115
Q1	0.6650	0.3783
Q2	0.6978	0.6373
mean	0.6957	0.5425
Q3	0.7283	0.6715
max	0.8364	0.7380

Method 2) u – space

Proposal density Sample Path	Unif(-1,1)	Unif(-0.01,0.01)
		
Histogram of $\delta^{(t)}$ after burn in the first 200		
min	0.5201	0.5889
Q1	0.6660	0.6475
Q2	0.6997	0.6947
mean	0.6976	0.6943
Q3	0.7295	0.7350
max	0.8441	0.7894

앞에서 언급했듯이 두 방법(δ – space & u – space) 모두 동일한 chain 의 양상을 보여준다. $\epsilon \sim \text{Unif}(-1,1)$ 일 때는 target density 에 잘 수렴하지만 $\epsilon \sim \text{Unif}(-0.01,0.01)$ 인 경우 초기값에서 빠르게 벗어나지 못하고 δ – space에서 움직이는 폭이 좁다.

HW3. 예제 7.5

A clinical trial was conducted to determine whether a hormone treatment benefits women who were treated previously for breast cancer. Each subject entered the clinical trial when she had a recurrence. She was then treated by irradiation and assigned to either a hormone therapy group or a control group. The observation of interest is the time until a second recurrence, which may be assumed to follow an exponential distribution with parameter $\tau\theta$ (hormone therapy group) or θ (control group). Many of the women did not have a second recurrence before the clinical trial was concluded, so that their recurrence times are censored.

TABLE 7.2 Breast cancer data.

	Hormone Treated								Control			
Recurrence	2	4	6	9	9	9	1	4	6	7	13	24
Times	13	14	18	23	31	32	25	35	35	39		
	33	34	43									
Censoring	10	14	14	16	17	18	1	1	3	4	5	8
Times	18	19	20	20	21	21	10	11	13	14	14	15
	23	24	29	29	30	30	17	19	20	22	24	24
	31	31	31	33	35	37	24	25	26	26	26	28
	40	41	42	42	44	46	29	29	32	35	38	39
	48	49	51	53	54	54	40	41	44	45	47	47
	55	56					47	50	50	51		

In Table 7.2, a censoring time M means that a woman was observed for M months and did not have a recurrence during that time period, so that her recurrence time is known to exceed M months. For example, 15 women who received the hormone treatment suffered recurrences, and the total of their recurrence times is 280 months.

Let $y_i^H = (x_i^H, \delta_i^H)$ be the data for the i th person in the hormone group, where x_i^H is the time and δ_i^H equals 1 if x_i^H is a recurrence time and 0 if a censored time. The data for the control group can be written similarly.

The likelihood is then

$$L(\theta, \tau | y) \propto \theta^{(\sum \delta_i^C + \sum \delta_i^H)} \tau^{(\sum \delta_i^H)} \exp\{-\theta \sum x_i^C - \tau \theta \sum x_i^H\}.$$

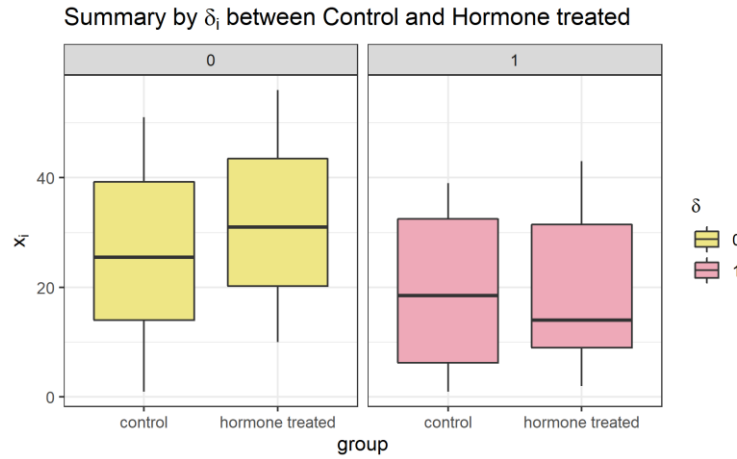
You've been hired by drug company to analyze their data. They want to know if the hormone treatment works, so the task is to find the marginal posterior distribution of τ using the Gibbs sampler. In a Bayesian analysis of these data, use the conjugate prior

$$f(\theta, \tau) \propto \theta^a \tau^b \exp\{-c\theta - d\tau\}$$

Physicians who have worked extensively with this hormone treatment have indicated that reasonable values for the hyperparameters are $(a, b, c, d) = (3, 1, 6, 20)$.

a. Summarize and plot the data as appropriate.

	Hormone Treated		Control	
recurrence	1	0	1	0
min	2	10	1	1
Q1	9	20.25	6.25	14
Q2	14	31	18.50	25.50
mean	18.67	32.79	18.90	26.10
Q3	31.50	43.50	32.50	39.25
max	43	56	39	51



Summary statistics 와 Boxplot 을 통해 유방암이 재발한 경우($\delta_i = 1$) 재발 발생 시점이 재발이 발생하지 않은 경우($\delta_i = 0$)의 시점보다 짧다는 것을 알 수 있다. 특히 유방암이 재발하지 않은 경우 hormone 처방은 받은 그룹이 control 그룹 보다 시간의 간격이 길다.

b. Derive the conditional distributions necessary to implement the Gibbs sampler.

$$\begin{aligned}
 f(\theta, \tau | y) &\propto f(\theta, \tau) \times L(\theta, \tau) \\
 &\propto \theta^a \tau^b \exp\{-c\theta - d\theta\tau\} \theta^{(\sum \delta_i^C + \sum \delta_i^H)} \tau^{(\sum \delta_i^H)} \exp\{-\theta \sum x_i^C - \tau \theta \sum x_i^H\} \\
 &\propto \theta^{(\sum \delta_i^C + \sum \delta_i^H + a)} \tau^{(\sum \delta_i^H + b)} \exp\{-c\theta - d\theta\tau - \theta \sum x_i^C - \tau \theta \sum x_i^H\}
 \end{aligned}$$

Where $f(\theta, \tau | y)$ is posterior of θ, τ
 $f(\theta, \tau)$ is prior of θ, τ and $L(\theta, \tau)$ is likelihood of θ, τ

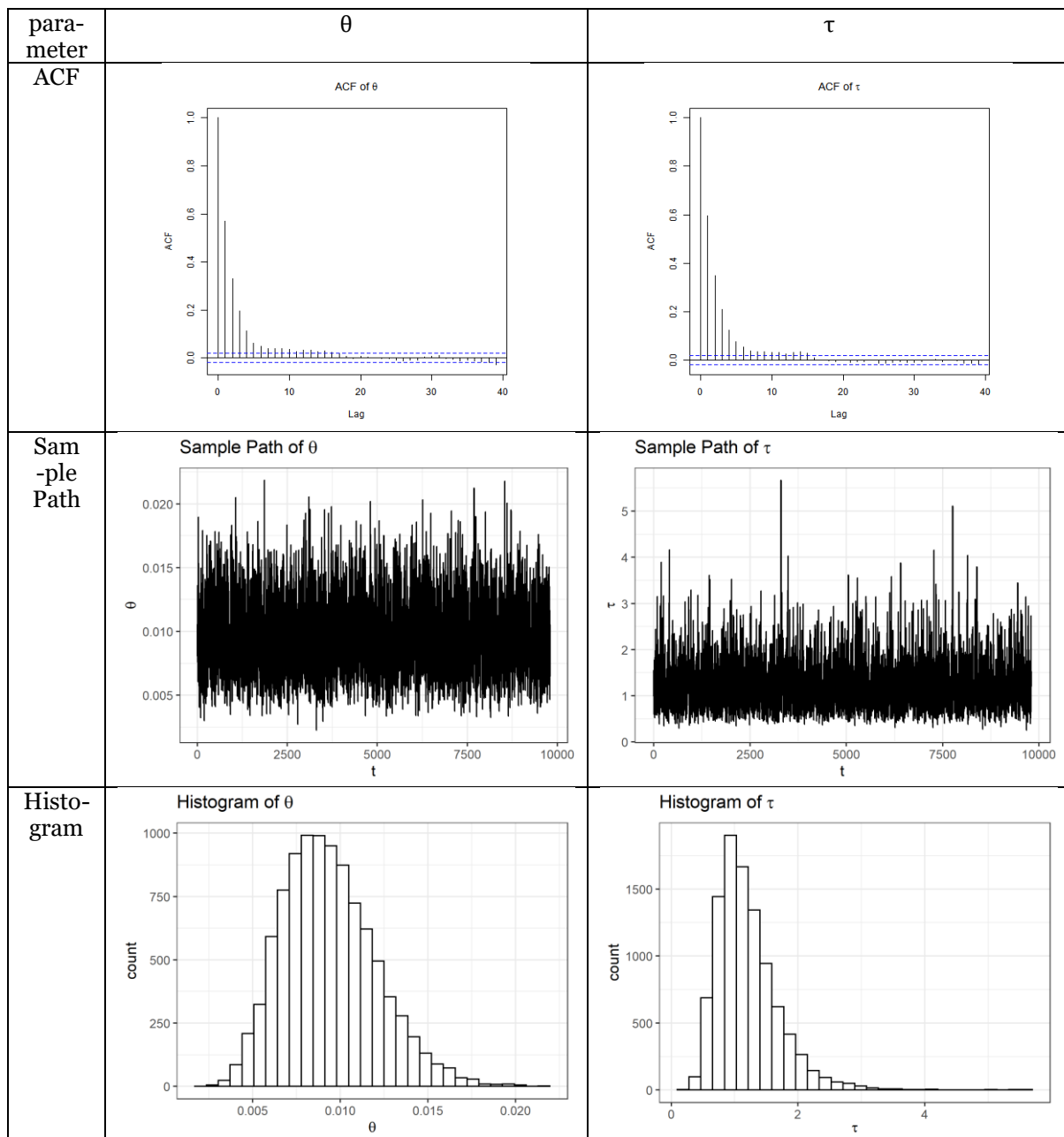
conditional distribution of θ given τ :

$$f(\theta | \tau) \sim \text{Gamma}(\sum \delta_i^C + \sum \delta_i^H + a + 1, c + \sum x_i^C + \tau(d + \sum x_i^H))$$

conditional distribution of τ given θ :

$$f(\tau | \theta) \sim \text{Gamma}(\sum \delta_i^H + b + 1, \theta(d + \sum x_i^H))$$

c. Program and run your Gibbs sampler. Use a suite of convergence diagnostics to evaluate the convergence and mixing of your sampler. Interpret the diagnostics.



위의 그래프는 Gibbs sampler 를 통해 10,000 개를 샘플링을 한 후 200 개를 burn in 결과를 ACF, Sample path, 그리고 Histogram 으로 보여준다. Sample 의 ACF 가 iteration 횟수가 증가함에 따라 급격히 감소하여 거의 0에 가까워진다. 샘플의 신뢰성을 높이기 위해서 thinning 이 필요한 것으로 보인다. Sample path 의 양상으로 보아 target density 에 도달한 것으로 보인다. 각 parameter 의 히스토그램을 보았을 때, θ 의 경우 0.01 근방에서 가장 많이 샘플링이 되었고 τ 의 경우 1 근방에서 가장 많이 샘플링 되었다.

d. Compute summary statistics of the estimated joint posterior distribution, including marginal means, standard deviations, and 95% probability intervals for each parameter. Make a table of these results.

	min	Q1	Q2	mean	Q3	max	sd	95% CI_L	95% CI_U
$f(\theta, \tau y)$	2.8e-73	4.5e-70	9.e-70	8.9e-70	1.3e-69	1.7e-69			
θ	0.0023	0.0074	0.0091	0.0094	0.0110	0.0219	0.0027	0.0041	0.0146
τ	0.2542	0.8630	1.1056	1.1978	1.4334	5.6697	0.4796	0.2578	2.1379

e. Create a graph which shows the prior and estimated posterior distribution for τ superimposed on the same scale.

Marginal prior of τ :

$$f(\tau) \propto \int_0^\infty f(\theta, \tau) d\theta \propto \int_0^\infty \theta^{(a+1)-1} e^{-\theta(c+d\tau)} \times \frac{(c+d\tau)^{a+1}}{\gamma(a+1)} d\theta \times \frac{\gamma(a+1)}{(c+d\tau)^{a+1}} \tau^b \propto \frac{\gamma(a+1)}{(c+d\tau)^{a+1}} \tau^b$$

$$(\because \int_0^\infty \theta^{(a+1)-1} e^{-\theta(c+d\tau)} \times \frac{(c+d\tau)^{a+1}}{\gamma(a+1)} d\theta = 1 \sim \text{Gamma}(a+1, c+d\tau))$$

Marginal posterior of τ :

$$f(\tau|y, \theta) \propto \int_0^\infty f(\theta, \tau|y) d\theta \propto \int_0^\infty \theta^{(\sum \delta_i^C + \sum \delta_i^H + a)-1} \tau^{(\sum \delta_i^H + b)} \exp\{-c\theta - d\theta\tau - \theta \sum x_i^C - \tau \theta \sum x_i^H\} d\theta$$

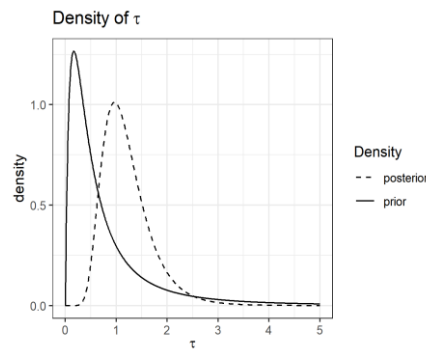
$$\propto \int_0^\infty \theta^{(\sum \delta_i^C + \sum \delta_i^H + a+1)-1} \exp\{-\theta(c + d\tau + \sum x_i^C + \tau \sum x_i^H)\} \times \frac{(c+d\tau + \sum x_i^C + \tau \sum x_i^H)^{\sum \delta_i^C + \sum \delta_i^H + a+1}}{\gamma(\sum \delta_i^C + \sum \delta_i^H + a+1)} d\theta$$

$$\times \frac{\gamma(\sum \delta_i^C + \sum \delta_i^H + a+1)}{(c+d\tau + \sum x_i^C + \tau \sum x_i^H)^{\sum \delta_i^C + \sum \delta_i^H + a+1}} \tau^{(\sum \delta_i^H + b)} \propto \frac{\gamma(\sum \delta_i^C + \sum \delta_i^H + a+1)}{(c+d\tau + \sum x_i^C + \tau \sum x_i^H)^{\sum \delta_i^C + \sum \delta_i^H + a+1}} \tau^{(\sum \delta_i^H + b)}$$

$$(\because \int_0^\infty \theta^{(\sum \delta_i^C + \sum \delta_i^H + a+1)-1} \exp\{-\theta(c + d\tau + \sum x_i^C + \tau \sum x_i^H)\} \times \frac{(c+d\tau + \sum x_i^C + \tau \sum x_i^H)^{\sum \delta_i^C + \sum \delta_i^H + a+1}}{\gamma(\sum \delta_i^C + \sum \delta_i^H + a+1)} d\theta = 1$$

$$\sim \text{Gamma}(\sum \delta_i^C + \sum \delta_i^H + a + 1, c + d\tau + \sum x_i^C + \tau \sum x_i^H))$$

아래의 그래프는 τ 의 prior density 와 posterior density 의 적분 값이 1 이 되도록 normalizing constant 를 구하여 겹쳐 그린 결과이다.



f. Interpret your results for the drug company. Specifically, what does your estimate of tau mean for the clinical trial? Are the recurrence time for the hormone group significantly different from those for control group?

$$H_0: \tau = 1 \text{ vs } H_1: \tau > 1$$

$\bar{\tau}$	T-statistic	p-value	95% CI_L	95% CI_U
1.1978	41.489	< 2.2e-16	1.2	Inf

따라서 p-value < 2.2e-16 이므로 유의수준 $\alpha = 0.05$ 에서 귀무가설 $\tau = 1$ 을 기각한다. 즉 hormone treatment 는 재발 시점에 영향을 준다고 판단할 수 있다.

3. Discussion

본 과제를 통해 Monte Carlo Markov Chain 의 전반적인 이론과 실습을 배울 수 있었다. Markov chain 은 state 가 이전의 state 에만 영향을 받는 수열을 말한다. Irreducible 하고 aperiodic 한 Markov chain 은 ergodic 한 성질을 가지며 이러한 chain 은 unique 한 stationary density 를 갖는다.

Unique 한 stationary density 를 갖는 Markov chain 을 생성하는 방법은 크게 두가지로 Metropolis-Hastings 와 Gibbs sampling 이다. Metropolis-Hastings 는 proposal density 인 g 에서 sample 을 뽑아 Metropolis-Hastings ratio 의 확률을 갖는 proposal 을 accept 하는 방식으로 chain 을 생성한다. 본 과제에서는 Metropolis-Hastings 를 응용하여 independence chain 과 random walk chain 을 생성하였다. 하지만 Metropolis-Hastings 는 proposal density 에 따라 chain 의 수렴여부가 달라진다는 단점이 있다. 한편 Gibbs Sampling 은 각 parameter 의 완전 조건부 확률에서 각각의 parameter 들을 순차적으로 샘플링 한다. 각 parameter 의 완전 조건부 확률의 유도가 간단한 경우 빈번히 사용되는 방법이다.

Bayesian inference 는 posterior 의 normalizing constant 를 구하는 데 어려움이 있어 직접적인 통계 추론의 제약을 갖는 경우가 많다. 하지만 posterior density 를 stationary density 로 unique 하게 갖는 Markov chain 을 생성한다면 normalizing constant 를 구하지 않아도 stationary density 에서 샘플링을 통해 통계 추론을 할 수 있기 때문에 MCMC 는 강력한 베이지안 통계 추론의 도구가 된다.

4. Appendix

```
pacman::p_load(ggplot2, tidyverse, dplyr, plotly, processx)
rm(list=ls(all=TRUE))
setwd('C:/Users/dnskd/Desktop/19-2/계통/과제/hw9')

## mixture model histogram
dat <- read.table('mixture.dat', header=T) # true delta 0.7
```

```

dat %>% ggplot(aes(y))+geom_histogram(aes(y=..density..),color="black",fill='white',bins=30)+
  theme_minimal()+theme_bw()+labs(x="y", y="Density", title="Histogram of Mixture data from EX
7.2")+geom_density(col='red')
ggsave('Histogram of mixture.png')

## mixture model
mixture <- function(delta){
  y <- dat$y
  prod(delta*dnorm(y, 7, 0.5) + (1-delta)*dnorm(y, 10, 0.5))
}

##HW1. MCMC Independence Chain
MCMC <- function(a=1 , b=1, n.sim=10^4){
  # initial value
  delta <- runif(1)

  for(i in 2:n.sim){
    current <- delta[i-1]
    cand <- rbeta(1, a, b)
    R <- mixture(cand)/mixture(current)
    delta[i] <- ifelse( rbinom(1, 1, min(R,1)) == 1 , cand, current)
  }

  return(delta)
}

## independence chain 1
delta1 <- MCMC()
delta1.df <- delta1 %>% data.frame(); colnames(delta1.df) <- "d"
delta1.df$ind <- 1:10000

## independence chain 2
delta2 <- MCMC(a=2, b=10)
delta2.df <- delta2 %>% data.frame(); colnames(delta2.df) <- "d"
delta2.df$ind <- 1:10000

## delta path plot
delta1.df %>% ggplot(aes(x=ind,y=d))+geom_line()+theme_minimal()+theme_bw()+
  labs(x="t", y=expression(delta), title="Sample Path for delta from Beta(1,1)")
ggsave('Delta1_path.png')

delta2.df %>% ggplot(aes(x=ind, y=d))+geom_line()+theme_minimal()+theme_bw()+
  labs(x="t", y=expression(delta), title="Sample Path for delta from Beta(2,10)")
ggsave('Delta2_path.png')

## delta histogram
delta1.df <- delta1.df[-c(1:200),]
delta1.df %>% ggplot(aes(d))+geom_histogram(bins=30, col='black',
fill='white')+theme_minimal()+theme_bw()+
  labs(x=expression(delta),title="Histogram of delta generated from Beta(1,1)")
ggsave('Delta1_Histogram.png')

delta2.df <- delta2.df[-c(1:200),]
delta2.df %>% ggplot(aes(d))+geom_histogram(bins=30, col='black',
fill='white')+theme_minimal()+theme_bw()+
  labs(x=expression(delta),title="Histogram of delta generated from Beta(2,10)")
ggsave('Delta2_Histogram.png')

## delta statistics

```

```

summary(delta1)
summary(delta2)

# HW 2 Random Walk Chain

# M1) Delta space
rm(list=ls())
df <- read.csv('mixture.dat')

## mixture model
mixture <- function(delta){
  y <- df$y
  prod(delta*dnorm(y, 7, 0.5) + (1-delta)*dnorm(y, 10, 0.5))
}

logit <- function(x) log(x/(1-x))
logitinv <- function(x) exp(x)/(1+exp(x))
Jacob <- function(x) 1/(x*(1-x))

rw.d <- function(b=1, n.sim=10^4){
  # initial value
  d <- runif(1)

  for(i in 2:n.sim){
    u <- logit(d[i-1])
    u.star <- u + runif(1, -b, b)
    d.star <- logitinv(u.star)

    MHR <- ( mixture(d.star) * dunif(logit(d[i-1]) - logit(d.star), -b, b ) * abs(Jacob(d[i-1])) ) / ( mixture(d[i-1]) * dunif(logit(d.star) - logit(d[i-1]), -b, b) *abs(Jacob(d.star)))
    R <- min(1, MHR)

    d[i] <- ifelse(rbinom(1,1,R)==1, d.star, d[i-1])
  }
  return(d)
}

## Random Walk chain 1
rw.d1 <- rw.d()
rw.d1.df <- rw.d1 %>% data.frame(); colnames(rw.d1.df) <- "d"
rw.d1.df$ind <- 1:10000

## Random Walk chain 2
rw.d2 <- rw.d(b=0.01)
rw.d2.df <- rw.d2 %>% data.frame(); colnames(rw.d2.df) <- "d"
rw.d2.df$ind <- 1:10000

## delta path plot
rw.d1.df %>% ggplot(aes(x=ind,y=d))+geom_line()+theme_minimal()+theme_bw()+
  labs(x="t", y=expression(delta), title="Sample Path for delta space when h ~ Unif(-1,1)")
ggsave('rw_d1_path.png')

rw.d2.df %>% ggplot(aes(x=ind, y=d))+geom_line()+theme_minimal()+theme_bw()+
  labs(x="t", y=expression(delta), title="Sample Path for delta space when h ~ Unif(-0.01,0.01)")
ggsave('rw_d2_path.png')

## delta histogram
rw.d1.df <- rw.d1.df[-c(1:200),]

```

```

rw.d1.df %>% ggplot(aes(d))+geom_histogram(bins=30, col='black',
fill='white')+theme_minimal()+theme_bw()+
  labs(x=expression(delta),title="Histogram of delta when h ~ Unif(-1,1)")
ggsave('rw_d1_Histogram.png')

rw.d2.df <- rw.d2.df[-c(1:200),]
rw.d2.df %>% ggplot(aes(d))+geom_histogram(bins=30, col='black',
fill='white')+theme_minimal()+theme_bw()+
  labs(x=expression(delta),title="Histogram of delta when h ~ Unif(-0.01,0.01)")
ggsave('rw_d2_Histogram.png')

## delta statistics
summary(rw.d1.df$d)
summary(rw.d2.df$d)

# M2) U-space
rm(list=ls())
df <- read.csv('mixture.dat')

## mixture model
mixture <- function(delta){
  y <- df$y
  prod(delta*dnorm(y, 7, 0.5) + (1-delta)*dnorm(y, 10, 0.5))
}

logitinv <- function(x) exp(x)/(1+exp(x))
logit <- function(x) log(x/(1-x))
Jacob <- function(x) exp(x)/(1+exp(x))^2

rw.u <- function(b=1, n.sim=10^4){
  # initial value
  u <- runif(1)
  d <- logitinv(u)

  for(i in 2:n.sim){
    u.star <- u + runif(1, -b, b)

    MHR <- ( mixture(logitinv(u.star)) * abs(Jacob(u.star)) * dunif(u - u.star, -b, b) ) /
    ( mixture(logitinv(u)) * abs(Jacob(u)) * dunif(u.star-u, -b, b))
    R <- min(1, MHR)

    u <- ifelse(rbinom(1,1,R)==1, u.star, u)
    d[i] <- logitinv(u)
  }
  return(d)
}

## Random Walk chain 1
rw.u1 <- rw.u()
rw.u1.df <- rw.u1 %>% data.frame(); colnames(rw.u1.df) <- "d"
rw.u1.df$ind <- 1:10000

## Random Walk chain 2
rw.u2 <- rw.u(b=0.01)
rw.u2.df <- rw.u2 %>% data.frame(); colnames(rw.u2.df) <- "d"
rw.u2.df$ind <- 1:10000

## delta path plot
rw.u1.df %>% ggplot(aes(x=ind,y=d))+geom_line()+theme_minimal()+theme_bw()+
  labs(x="t", y=expression(delta), title="Sample Path for u space when h ~ Unif(-1,1)")
ggsave('rw_u1_path.png')

```



```

rw.u2.df %>% ggplot(aes(x=ind, y=d))+geom_line()+theme_minimal()+theme_bw()+
  labs(x="t", y=expression(delta), title="Sample Path for u space when h ~ Unif(-0.01,0.01)")
ggsave('rw_u2_path.png')

## delta histogram
rw.u1.df <- rw.u1.df[-c(1:200),]
rw.u1.df %>% ggplot(aes(d))+geom_histogram(bins=30, col='black',
fill='white')+theme_minimal()+theme_bw()+
  labs(x=expression(delta),title="Histogram of delta in u space when h ~ Unif(-1,1)")
ggsave('rw_u1_Histogram.png')

rw.u2.df <- rw.u2.df[-c(1:200),]
rw.u2.df %>% ggplot(aes(d))+geom_histogram(bins=30, col='black',
fill='white')+theme_minimal()+theme_bw()+
  labs(x=expression(delta),title="Histogram of delta in u space when h ~ Unif(-0.01,0.01)")
ggsave('rw_u2_Histogram.png')

## delta statistics
summary(rw.u1.df$d)
summary(rw.u2.df$d)

##### HW3 #####
rm(list=ls())
hormone <- data.frame(
  x = c(2, 4, 6, 9, 9, 9,
        13, 14, 18, 23, 31, 32,
        33, 34, 43,
        10, 14, 14, 16, 17, 18,
        18, 19, 20, 20, 21, 21,
        23, 24, 29, 29, 30, 30,
        31, 31, 31, 33, 35, 37,
        40, 41, 42, 42, 44, 46,
        48, 49, 51, 53, 54, 54,
        55, 56),
  d = as.factor(c( rep(1, 15), rep(0, 38)))
)

control <- data.frame(
  x = c(1, 4, 6, 7, 13, 24,
        25, 35, 35, 39,
        1, 1, 3, 4, 5, 8,
        10, 11, 13, 14, 14, 15,
        17, 19, 20, 22, 24, 24,
        24, 25, 26, 26, 26, 28,
        29, 29, 32, 35, 38, 39,
        40, 41, 44, 45, 47, 47,
        47, 50, 50, 51),
  d = as.factor(c( rep(1, 10), rep(0, 40)))
)

# summary stat
summary(hormone[which(hormone$d==1),])
summary(hormone[which(hormone$d==0),])

summary(control[which(control$d==1),])
summary(control[which(control$d==0),])

```

```

# Box plot
df <- rbind(control, hormone)
df$group <- rep(c("control", "hormone treated"), c(nrow(control), nrow(hormone)))
df %>% ggplot(aes(group, x, fill=d)) + geom_boxplot() + facet_wrap(~d) +
labs(title=expression(paste("Summary by ", delta[i], " between Control and Hormone treated")),

y=expression(x[i]))+theme_minimal()+theme_bw()+scale_fill_manual(expression(delta),
values=c("khaki2", "pink2"))
ggsave('7.5.a.png')

# c. Gibbs sampler

##### start Gibbs #####

# initial values
n.sim <- 10^4
burn <- 200
theta <- runif(1)
tau <- runif(1)
sum.c <- sum(as.numeric(control$d)-1)
sum.h <- sum(as.numeric(hormone$d)-1)
a <- 3; b <- 1; c<- 60; d <- 120
sum.xc <- sum(control$x)
sum.xh <- sum(hormone$x)

# generate parameter
for( i in 2:n.sim ){
  theta[i] <- rgamma( 1, sum.c + sum.h + a + 1, c + sum.xc + tau[i-1] * ( d + sum.xh ) )
  tau[i] <- rgamma(1, sum.h + b + 1, theta[i] * (d + sum.xh))
}

# burn-in
theta <- theta[-c(1:burn)]
tau <- tau[-c(1:burn)]

##### end Gibbs #####

## acf graph
png("acf_theta.png"); acf(theta, main=expression(paste("ACF of ", theta))); dev.off()
png("acf_tau.png"); acf(tau, main=expression(paste("ACF of ", tau))); dev.off()

## sample path and histogram
theta.df <- theta %>% data.frame(); colnames(theta.df) <- "theta"; theta.df$t <-
1:length(theta)
theta.df %>% ggplot(aes(t, theta)) + geom_line() + theme_minimal() + theme_bw() +
labs(title=expression(paste("Sample Path of ", theta))),

y=expression(theta), x="t")
ggsave('Sample path of theta.png')
theta.df %>% ggplot(aes(theta)) + geom_histogram(fill="white", col='black', bins=30) +
labs(title=expression(paste("Histogram of ", theta))),

x=expression(theta) + theme_minimal() + theme_bw()
ggsave('Histogram of theta.png')

tau.df <- tau %>% data.frame(); colnames(tau.df) <- "tau"; tau.df$t <- 1:length(tau)
tau.df %>% ggplot(aes(t, tau)) + geom_line() + theme_minimal() + theme_bw() +
labs(title=expression(paste("Sample Path of ", tau))),

```

```

y=expression(tau), x="t")
ggsave('Sample path of tau.png')
tau.df %>% ggplot(aes(tau)) + geom_histogram(fill="white", col='black', bins=30) +
labs(title=expression(paste("Histogram of ", tau))),

x=expression(tau)) + theme_minimal() + theme_bw()
ggsave('Histogram of tau.png')

## d.
prior <- function(theta, tau, a, b, c, d){
  theta^a * tau^b * exp(-c*theta -d*theta*tau)
}
likli <- function(theta, tau, a, b, c, d){
  theta^(sum.c + sum.h)*tau^(sum.h)*exp(-theta*sum.xc-tau*theta*sum.xh)
}
post <- function(theta, tau, a, b, c, d){
  prior(theta, tau, a, b, c, d) * likli(theta, tau, a, b, c, d)
}

post.density <- post(theta=theta, tau=tau, a=3, b=1, c=60, d=120)
summary(post.density)
summary(theta); summary(tau)
sd(theta); sd(tau)
mean(theta) - qnorm(0.975) * sd(theta)
mean(theta) + qnorm(0.975) * sd(theta)

mean(tau) - qnorm(0.975) * sd(tau)
mean(tau) + qnorm(0.975) * sd(tau)

## e. tau prior and posterior density
tau.prior <- function(tau, a, b, c, d){
  return((gamma(a+1) * tau^b)/(c+d*tau)^(a+1))
}
tau.post <- function(tau, a, b, c, d){
  return((gamma(sum.c+sum.h+a+1) * tau^(sum.h+b))/(c+d*tau+sum.xc+tau*sum.xh)^(sum.c+sum.h+a+1))
}
c1 <- 1/integrate(tau.prior, a=3, b=1, c=60, d=120, lower=0, upper=Inf)$value
c2 <- 1/integrate(tau.post, a=3, b=1, c=60, d=120, lower=0, upper=Inf)$value

tau.p <- data.frame(tau = seq(0, 5, length=10^3))
tau.p$prior <- c1*tau.prior(tau=tau.p$tau,a=3, b=1, c=60, d=120)
tau.p$post <- c2*tau.post(tau=tau.p$tau,a=3, b=1, c=60, d=120)
tau.p %>% ggplot()+geom_line(aes(tau,prior,linetype="prior"))+geom_line(aes(tau,
post,linetype="posterior"))+
  theme_minimal()+theme_bw()+labs(title=expression(paste("Density of ", tau)),
x=expression(tau), y="density")+
  scale_linetype_manual("Density", values=c("prior"=1, "posterior"=2))
ggsave('densityoftau.png')

## f. hypothesis test

t.test(tau, mu=1, alternative="greater",conf.level=0.95)

```