Computational Statistics

HW#9

182STG18 이하경

Ⅰ. Description

MCMC (Markov Chain Monte-Carlo Simulation)

Markov-Chain

$$\text{a sequence of random variables } \{X^{(t)}\}, \qquad t = 0, 1, \dots$$

$$\text{state } X^{(t)} = \text{ a finite or countably infinite number of values } \ \& \ \text{ state space } \mathcal{S} = \text{the set of possible values of } X^{(t)}$$

위와 같은 확률변수들의 배열에서 이들의 결합확률은 아래와 같이 각 확률변수들의 조건부 확률의 곱으로 표현할 수 있고, $X^{(t)}$ 들이 서로 조건부 독립이라면 아래와 같은 Markov Property, one-step memory를 만족한다.

$$P[X^{(0)}, \dots, X^{(n)}] = P[X^{(n)} \mid x^{(0)}, \dots, x^{(n-1)}] \cdot P[X^{(n-1)} \mid x^{(0)}, \dots, x^{(n-2)}] \cdot \cdots \cdot P[X^{(1)} \mid x^{(0)}] \cdot P[X^{(0)}]$$

$$= P[X^{(n)} \mid x^{(n-1)}] \cdot P[X^{(n-1)} \mid x^{(n-2)}] \cdot \cdots \cdot P[X^{(1)} \mid x^{(0)}] \cdot P[X^{(0)}]$$

모든 t에 대해 $p_{ij}^{(t)} = P[X^{(t+1)} = j \mid X^{(t)} = i]$ 을 만족할 때 이러한 sequence $\{X^{(t)}\}$ 을 Markov Chain이라고 하며, $p_{ij}^{(t)}$ 을 one-step transition probability라고 한다. 다시 말해 t 시점의 발생확률이 바로 이전인 t-1 시점에 따른 조건부확률에만 의존하며 이전의 상황은 영향을 미치지 않는다는 것을 의미한다.

만약 어떤 Markov Chain의 trainsition probability가 모든 $i, j \in \mathcal{S}$와 stationary distribution π에 대해 $\pi_i p_{ij} = \pi_j p_{ji}$ 을 만족할 경우 chain이 reversible하다고 하며, π가 irreducible, aperiodic 성질을 만족하면 t가 커짐에 따라 각각의 transition probability P_j는 π_j로 수렴한다.

$$\lim_{n \to \infty} P[X^{(x+n)} = j \mid X^{(t)} = i] = \pi_j$$

만약 $X^{(1)}, X^{(2)}, \dots$ 이 stationary distribution π를 가지는 irreducible & aperiodic Markov chain으로부터 온 값들이라고 할 때 임의의 함수 h에 대해, 대수의 법칙(LLN)에 의한 다음의 성질을 만족한다.

$$\frac{1}{n}\sum_{t=1}^{n} h(X^{(t)}) \to E_\pi\{h(X)\} \quad as \quad n \to \infty$$

따라서 Markov Chain을 MC simulation에 적용하여 관심 기댓값을 추정할 수 있다. 즉 MCMC란 irreducible & aperiodic chain을 적절히 설정함으로써 simulation을 반복하여 얻어지는 sample의 수렴하는 stationary distribution이 실제 target distribution f와 같아지기를 기대하는 simulation 방법이라고 할 수 있다.

1. MetroPolis-Hasting Algorithm

M-H 알고리즘은 Markov Chain을 설정하기 위한 매우 일반적인 방법이다.

0. 초기 분포 g로부터 $X^{(0)} = x^{(0)}$을 임의 추출한다.
1. 다음 후보가 되는 X^*을 proposal distribution $g(\cdot \mid x^{(0)})$로부터 임의로 추출한다.
2. M-H Ratio $R(X^*, x^{(0)}) = \frac{f(X^*)\, g(x^{(0)} \mid X^*)}{f(x^{(0)})\, g(X^* \mid x^{(0)})}$을 계산한다.
3. min(1, R), 1-min(1, R)의 확률로 δ^*와 $\delta^{(0)}$ 중 하나를 임의로 추출하여 $X^{(1)}$으로 지정한다.

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

4. 1-3을 n번 반복하여 총 크기 n의 X의 sample을 얻는다.

여기서 proposal distribution이 $g(x^{(t)}|x^*) = g(x^*|x^{(t)})$으로 symmetric하다면 Metropolis Algorithm이며, chain은 Markov property를 가진다. chain을 거쳐 처음 stationary distribution에 도달할 경우 그 다음의 모든 단계에서 추출되는 값들은 모두 같은 marginal distribution을 가진다. 따라서 위에서 말했듯이 sample을 draw할 proposal distiribution g를 적절히 선택하여 target distribution으로 수렴하도록 하는 것이 관건이다.

1.1 Independence Chains

M-H algorithm의 proposal distribution g가 $g(x^*|x^{(t)}) = g(x^*)$을 만족하는 independent chain으로 각 단계의 candidate sample이 이전 단계와 서로 독립을 이룬다고 가정한 후 sampling하는 간단한 방법이다.

1.2 Random Walk Chains

Random Walk Chain은 Markov Chain의 또 다른 유형으로 특정 density h로부터 생성된 $\epsilon \sim h(\epsilon)$에 대해 $X^* = x^{(t)} + \epsilon$의 관계를 설정하여 이전 단계의 값에 일종의 noise를 추가하여 다음 단계의 후보 값을 draw하는 방법이라고 할 수 있다. 일반적으로 h 분포는 Uniform, scaled standard normal, scaled Student's t 분포를 이용한다.

Implementation 1에서는 Markov Chain을 Independent Chain과 Random Walk Chain으로 각각 설정하여 GMM의 mixing proportion δ의 target distribution을 찾아본다.

2. Gibbs Sampling

위에서 모두 1차원의 sampling을 고려했다면, Gibbs Sampling은 multidimensional target distribution에 적용 가능한 매우 효과적인 sampling 방법이다. $X = (X_1, \dots, X_p)^T, X_{-i} = (X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_p)^T$라고 할 때 $X_i \mid X_{-i} = x_{-i}$으로 i번째 variable을 제외한 다른 값들이 주어진 상황에서 Xi의 univariate conditional density $f(x_i|x_{-i})$는 다음의 Gibbs sampling 방법으로 생성할 수 있다.

1. 초기값 $x^{(0)} = \left(x_1^{(0)}, \dots, x_p^{(0)}\right)^T$을 설정한다.
2. 다음에 따라 각 변수의 값을 단계를 거쳐 최신 값으로 update한다.

$$
\begin{aligned}
&X_1^{(t+1)} | \cdot \sim f\left(x_1 \middle| x_2^{(t)}, \dots, x_p^{(t)}\right), \\
&X_2^{(t+1)} | \cdot \sim f\left(x_2 \middle| x_1^{(t+1)}, x_3^{(t)}, \dots, x_p^{(t)}\right), \\
&\quad \vdots \\
&X_p^{(t+1)} | \cdot \sim f(x_p | x_1^{(t+1)}, \dots, x_{p-1}^{(t+1)})
\end{aligned}
$$

3. 1-2를 draw할 sample의 크기만큼 반복한다.

Implementation 2에서는 유방암 재발 여성에 대한 호르몬 치료 효과 여부를 실험하기 위해 Gibbs sampling을 이용해 관심 모수의 target marginal distribution을 찾아본다.

II. Implementation

1. Mixture Distribution (Example 7.2)

Goal

$$y_1, \dots, y_{100} \sim iid\ Mixture\ Dist^n = \delta N(7, 0.5^2) + (1-\delta)N(10, 0.5^2),$$
$$data\ were\ generated\ from\ mixing\ proportion\ \delta = 0.07$$

MCMC Simulation에서 Independent Chain과 Random Walk chain을 적절히 설정하여 δ의 stationary distribution이 target distribution인 posterior와 같아지도록 하려고 한다. 이를 위해 적절한 proposal distribution $g(\cdot\,|\delta^{(0)})$을 가정한다.

0. 초기값 $\delta^{(0)}$을 초기 분포 U(0, 1)로부터 임의로 추출한다.
1. 다음 후보가 되는 δ^*을 proposal distribution $g(\cdot\,|\delta^{(0)})$으로부터 임의로 추출한다.
2. M-H Ratio $R(\delta^*, \delta^{(0)}) = \frac{f(\delta^*)\, g(\delta^{(0)}|\delta^*)}{f(\delta^{(0)})\, g(\delta^*|\delta^{(0)})}$ 을 계산한다.
3. min(1, R), 1-min(1, R)의 확률로 δ^*와 $\delta^{(0)}$ 중 하나를 임의로 추출하여 $\delta^{(1)}$으로 지정한다.
4. 1-3을 n번 반복하여 총 크기 n의 δ sample을 얻는다.

1) Independent Chain에서 δ의 제안 분포 $g(\cdot\,|\delta^{(t)}) = g(\cdot)$으로 $\mathrm{Beta}(1,1)$, $\mathrm{Beta}(2,10)$의 두 가지 경우를 고려한다.

$$\frac{f(\delta^*)\, g(\delta^{(t)})}{f(\delta^{(t)})\, g(\delta^*)} = \frac{L(\delta^*|y)}{L(\delta^{(t)}|y)}$$

2) Random Walk Chain의 $\delta^* = \delta^{(t)} + \epsilon$에서 ϵ의 분포로 $\mathrm{Unif}(-b, b)$을 가정할 때 b=1, b=0.01의 두 가지 경우를 고려한다. 이 경우 mixing proportion δ는 항상 0과 1사이의 값을 가져야 하므로 $\mathrm{logit}\{\delta\} = \log\left(\frac{\delta}{1-\delta}\right) = u$의 변환을 거쳐 u-space에서 $u^* = u^{(t)} + \epsilon,\ \epsilon \sim Unif(-b, b)$을 이용해 sample을 생성한 뒤 다시 역변환을 거쳐 최종 δ의 sample을 얻는다.

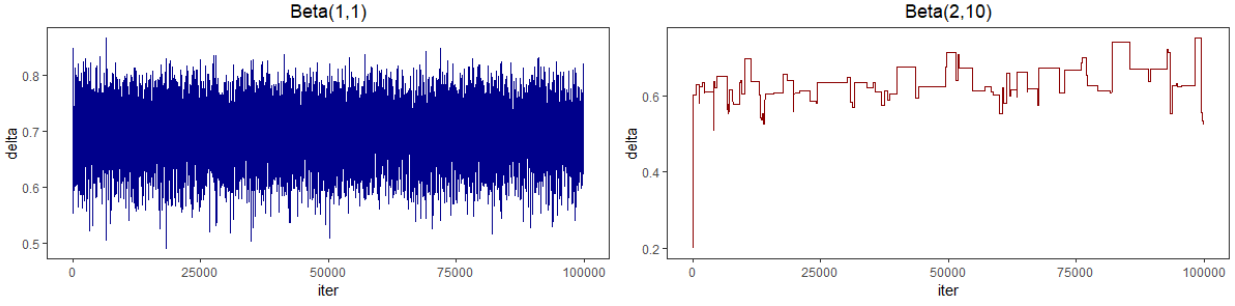
$$\frac{f(\delta^*)\, g(\delta^{(t)}|\delta^*)}{f(\delta^{(t)})\, g(\delta^*|\delta^{(0)})} = \frac{f(logit^{-1}\{u^*\})|J(u^*)|g(u^{(t)}|u^*)}{f(logit^{-1}\{u^{(t)}\})|J(u^{(t)})|g(u^*|u^{(t)})} = \frac{L(logit^{-1}\{u^*\}|y)}{L(logit^{-1}\{u^{(t)}\}|y)} \cdot \frac{|J(u^*)|}{|J(u^{(t)})|}$$

각각의 Chain에서 n=100,000의 δ의 sample을 생성하여 그래프(sample path 및 histogram)와 평균, 표준편차 등의 기술 통계량의 결과를 통해 사전분포 및 ϵ의 분포 설정에 따른 결과를 비교한다.

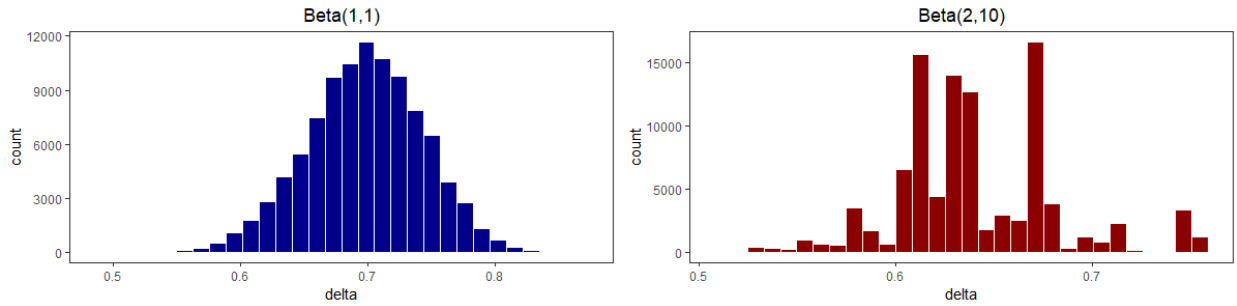
n=100,000개의 sample $\delta^{(t)}$에서 초기 500개의 δ는 제외하고 최종 99,500개의 sample을 생성하였다.

1) Independent Chains

Result 1-1. Sample Paths for $\delta^{(t)}$, $t = 1, \dots, 100,000$



Result 1-2. Histograms of $\delta^{(t)}$ for iterations 501-100,000



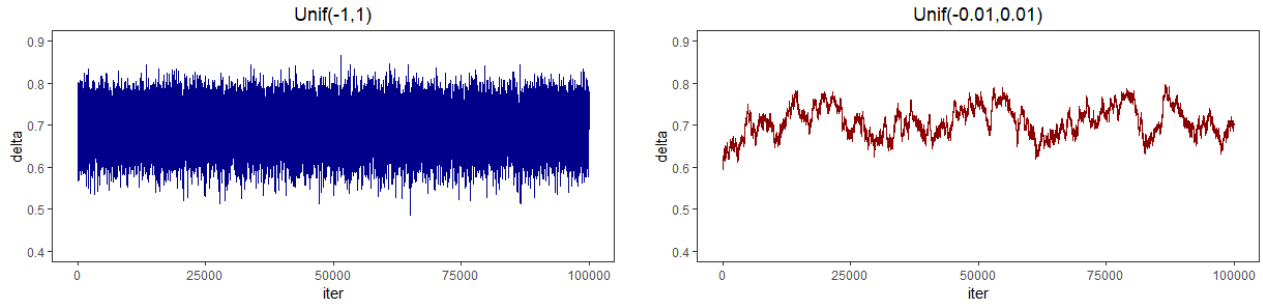
Result 1-3. Summary Statistics of $\delta^{(t)}$

prior	mean	sd	min	Q1	Q2	Q3	max	95% CI	
Beta(1,1)	**0.69656**	0.04546	0.44871	0.66629	0.69885	0.72753	0.86181	**0.60268**	**0.78071**
Beta(2,10)	**0.62419**	0.04256	0.52191	0.59951	0.62689	0.63535	0.74602	**0.55495**	**0.69546**

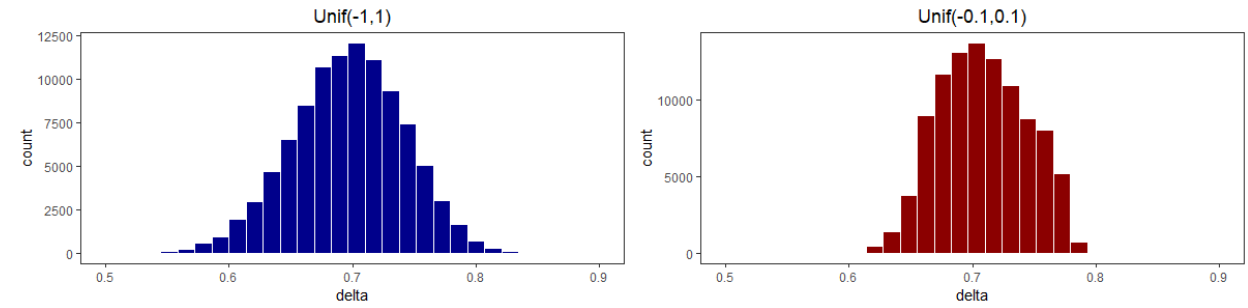
제시된 두 가지 제안 분포에 따라 결과가 매우 큰 차이를 보이고 있다. Beta(1, 1), 즉 U(0, 1)에서는 100,000번의 반복 동안 δ가 일정 범위 내에서 랜덤하게 추출되었으나 Beta(2, 10)에서는 일부구간에서 몇 차례의 반복 동안 δ가 update되지 않고 같은 값을 유지하는 것을 확인할 수 있다. 히스토그램에서도 Beta(2, 10)의 경우에는 뽑힌 sample의 δ들이 적절한 분포를 이루고 있다고 할 수 없고 Beta(1, 1)에 비해 좁은 범위에 형성되어 있으며, 평균 및 편차는 Beta(1, 1)에 비해 작았다. 이는 Beta(2, 10)을 제안분포로 하였을 경우 분산이 약 0.01로 매우 작아 각 반복 단계마다 후보가 되는 δ^*의 매우 좁은 범위 내에서 추출되어, 선택될 확률이 적기 때문이다. 따라서 Beta(2, 10)은 Markov Chain의 적절한 proposal distribution이라고 할 수 없다.

2) Random Walk Chains

Result 2-1. Sample Paths for $\delta^{(t)}$



Result 2-2. Histograms of $\delta^{(t)}$ for iterations 501-100,000



Result 1-3. Summary Statistics of $\delta^{(t)}$

$\epsilon \sim U(-b, b)$	mean	sd	min	Q1	Q2	Q3	max	95% CI	
b=1	**0.69641**	**0.04542**	0.48555	0.66628	0.69778	0.72799	0.86701	**0.60426**	**0.78168**
b=0.01	**0.70833**	**0.03537**	0.61168	0.68121	0.70677	0.73492	0.79685	**0.64501**	**0.77204**

b=0.01로 설정하였을 경우 다음 step의 candidate가 이전 step의 값에 $\epsilon \sim U(-0.01, 0.01)$만큼 추가되지만 이 값의 분산이 0을 둘레로 0.00003 정도로 매우 작아 target sample δ가 매우 좁은 범위에서만 형성되는 것을 볼 수 있다. 따라서 b=1의 경우 좋은 예가 되지만 b=0.01의 경우는 적절한 proposal distribution이 설정되지 못한다.

2. Clinical trial for breast cancer (Problems 7.5)

Goal

유방암 이력이 있는 여성에 대한 호르몬 치료의 효과에 대한 임상 시험이 시행되었다. 각각의 실험 대상 여성은 1차 재발 시점부터 실험에 투입된 후 호르몬 치료군과 대조군 중 하나로 분류되어, 2차 재발이 일어날 경우 해당 시점까지 걸리는 시간을 측정한다. 호르몬 치료군과 대조군의 재발 시점이 따르는 분포가 각각 $\mathrm{Exp}(\tau\theta)$, $\mathrm{Exp}(\theta)$이라고 가정할 때, 두 집단의 재발 시점에 차이가 있는지 평가하여 호르몬 치료의 효과에 대해 확인하고자 한다.

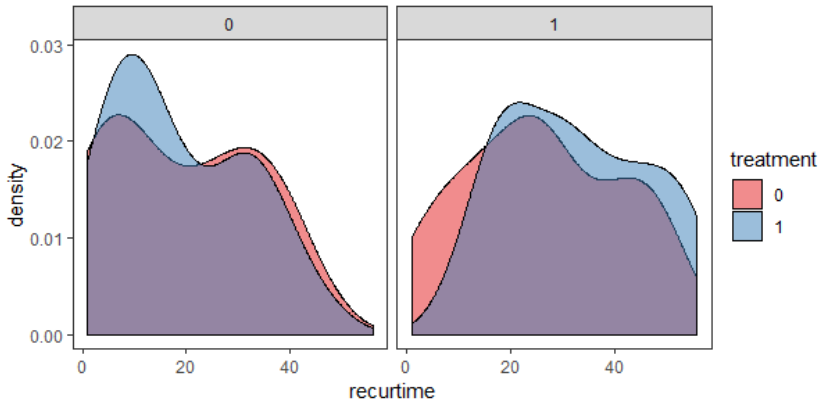
관측된 data와 제시된 conjugate prior, hyperparameter에 대한 τ의 marginal distribution을 찾기 위해, 적절한 조건부 분포를 설정한 Gibbs sampling을 시행한다. target distribution인 posterior로부터의 τ의 sample을 생성하고 평균과 예측 구간, 분포의 graph 등을 이용해 호르몬 치료 효과에 대해 검정하고자 한다.

Conjugate Prior	$f(\theta, \tau)$	$\theta^a \tau^b \exp\{-c\theta - d\theta\tau\}$
Likelihood	$L(\theta, \tau \mid y)$	$\theta^{\sum\delta^C + \sum\delta^H} \cdot \tau^{\sum\delta^H} \cdot \exp\{-\theta\sum x^C - \tau\theta\sum x^H\}$
Target Posterior	$f(\theta, \tau \mid y)$ $\propto f(\theta, \tau) \cdot L(\theta, \tau \mid y)$	$\theta^{a+\sum\delta^C+\sum\delta^H} \cdot \tau^{b+\sum\delta^H} \cdot \exp\{-c\theta - d\theta\tau - \theta\sum x^C - \tau\theta\sum x^H\}$

Gibbs sampling에서 $t = 1, \ldots, 100{,}000$에 대한 θ와 τ의 sample을 생성하고 초기 500개의 burn-in period를 제외한 총 99,500개의 sample을 구하였다.

a. Summary &Plot for the observed data

Group	Censored	min	Q1	Q2	mean	Q3	max	nobs
Hormone	O	10	20.25	31	32.79	43.5	56	38
	X	2	9	14	18.67	31.5	43	15
Control	O	1	14	25.5	26.1	39.25	51	40
	X	1	6.25	18.5	18.9	32.5	39	10



왼쪽 그래프는 실험 기간 중 재발이 일어난 recurrence time에 대해, 오른쪽 그래프는 재발이 일어나지 않고 실험 기간이 종료된 censored time의 분포를 호르몬 치료 여부에 따라 겹쳐 그린 것이다. 관측치의 평균과 분포로 보아 recurrence time의 경우 호르몬 치료 그룹이 조금 더 짧은 경향이 있으나 censored time의 경우 호르몬 치료 그룹이 더 길어 주어진 자료만으로 보아서는 호르몬 치료 효과에 대해 예상하기 어렵다.

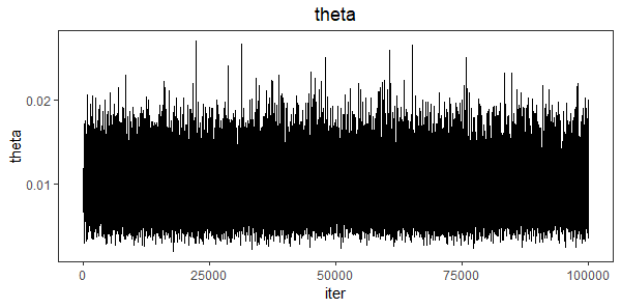
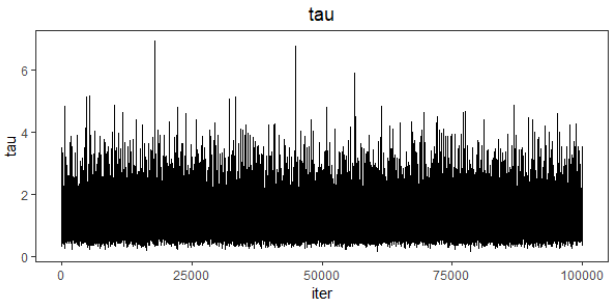
b. Derivation of the conditional distributions to implement the Gibbs sampler

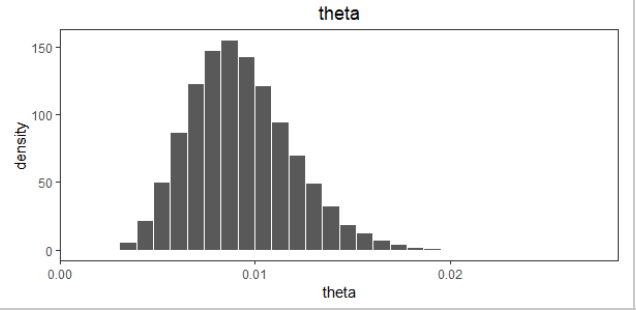
proposal density of θ	$g(\theta^{(t)} \mid \tau^{(t-1)},\ a,b,c,d) \propto \theta^{(t)\{a+\sum\delta^C+\sum\delta^H+1\}-1} \cdot \exp\{-(c+\sum x^C+d+\sum x^H \cdot \tau^{(t-1)})\,\theta^{(t)}\}$ $\propto \text{Gamma}(a+\sum\delta^C+\sum\delta^H+1,\ c+\sum x^C+d+\sum x^H \cdot \tau^{(t-1)})$
proposal density of τ	$g(\tau^{(t)} \mid \theta^{(t)},\ a,b,c,d) \propto \tau^{(t)\{b+\sum\delta^H+1\}-1} \cdot \exp\{-(d+\sum x^H \cdot \theta^{(t)})\,\tau^{(t)}\}$ $\propto \text{Gamma}(b+\sum\delta^H+1,\ d+\sum x^H \cdot \theta^{(t)})$

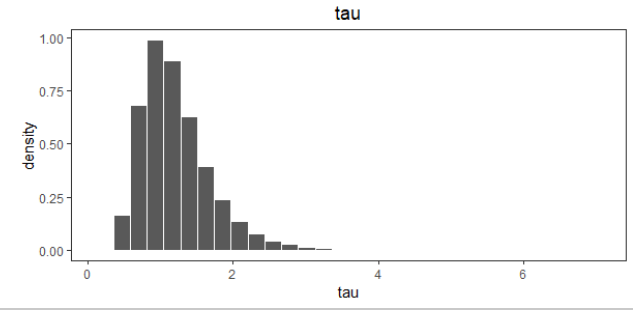
cf. $f(x) \propto x^{\alpha-1}\exp(-\lambda x) \sim Gamma(\alpha,\lambda)$

Target distribution인 posterior 분포에서 비례상수를 제외한 식을 θ와 τ가 각각 주어진 값일 때의 조건부 분포의 형태로 본다면 위와 같이 Gamma 분포에 비례하도록 표현할 수 있다. 따라서 전 단계의 $\tau^{(t-1)}$이 주어졌을 때 $\theta^{(t)}$를 shape parameter $a+\sum\delta^C+\sum\delta^H+1$, scale parameter $\left(c+\sum x^C+d+\sum x^H \cdot \tau^{(t-1)}\right)^{-1}$인 Gamma 분포로부터 추출하고, 새로운 $\theta^{(t)}$가 주어졌을 때 $\tau^{(t)}$를 shape parameter $b+\sum\delta^H+1$, scale parameter $\left(d+\sum x^H \cdot \theta^{(t)}\right)^{-1}$인 Gamma 분포로부터 새롭게 추출하여 값을 반복적으로 update한다.

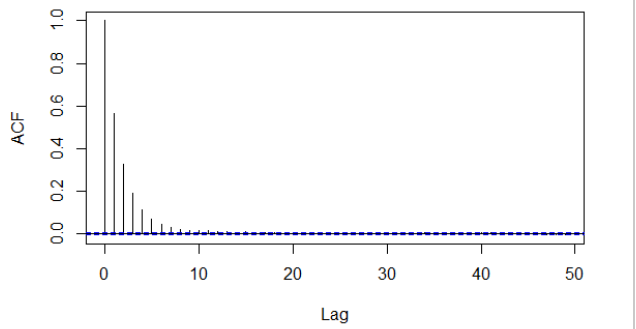
c. Run & Evaluation of the sampler using Convergence Diagnostics

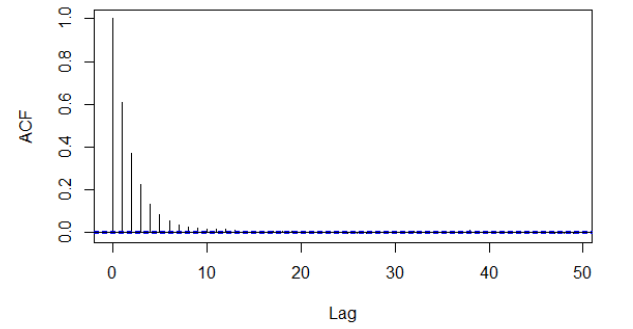
θ	τ
Sample paths	
	
Histogram	





Auto Correlation Function (ACF)



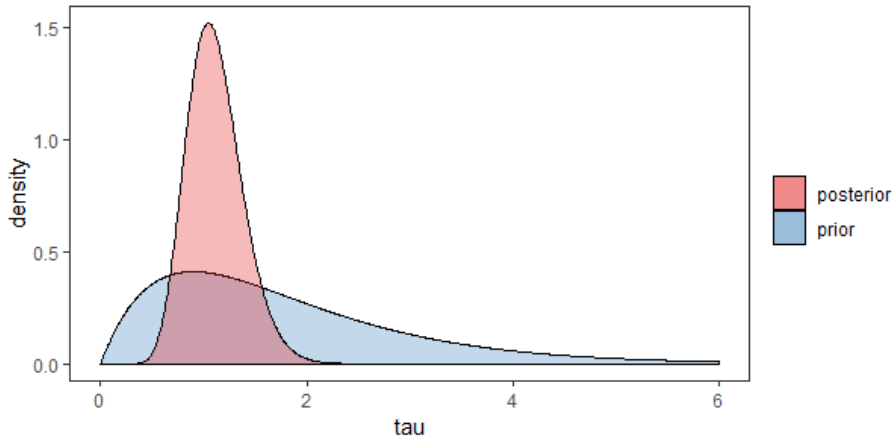


95,000번의 iteration 동안 θ와 τ은 각각 약 0에서 0.02, 0에서 4를 둘레로 랜덤하게 뽑혔으며 가끔씩 비교적 큰 값들이 포함된 것이 보인다. 히스토그램에서 확인할 수 있듯이 두 분포는 오른쪽 꼬리가 더 긴 형태를 보이고 있다. 자기상관함수 그래프에서 correlation이 빠르게 줄어드는 것을 보아 Gibbs sampling이 효율적으로 진행되었다고(good mixing) 할 수 있다.

d. Summary Statistics of the estimated joint posterior

	mean	sd	min	Q1	Q2	Q3	max	95% Prob.Interval	
θ	0.00928	0.00268	0.00190	0.00737	0.00902	0.01091	0.02705	0.00478	0.01526
τ	**1.21321**	0.49352	0.18500	0.86910	1.11960	1.45190	6.90970	**0.53531**	**2.43426**

e. Graph of Prior and Posterior distribution of τ



가정한 prior와 conditional distribution에 맞게 비례상수를 조정하여 tau의 범위에 따른 두 가지 density의 값을 구하고 그래프를 겹쳐 그려보았다. 그래프에서 왼쪽의 density는 τ 의 사전분포, 오른쪽 density는 사후분포를 나타낸다. hyperparameters (a, b, c, d)가 (3, 1, 60, 120)일 때 observed data에 기반한 사후분포는 Gamma(17, 15.27^{-1})으로 사전분포 (2, 1.11^{-1})에 비해 분산이 더 작다.

f. Interpretation of the meaning of τ for the clinical trial

$$H_0: \tau = 1 \quad vs \quad H_1: \tau \neq 1$$

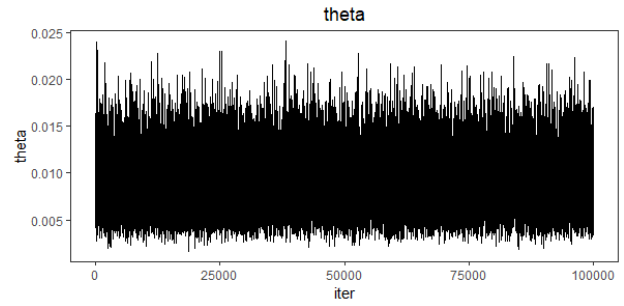
Result d에서 τ의 sample mean은 약 1.21으로, 95% Probability Interval (0.53509,2.42716)이 1을 포함하고 있다. 따라서 위와 같은 가설을 설정할 때 유의수준 $\alpha = 0.05$에서 귀무가설을 기각할 수 없다. 다시 말해, hormone therapy group과 control group의 recurrence time이 크게 다르다고 할 수 없다.

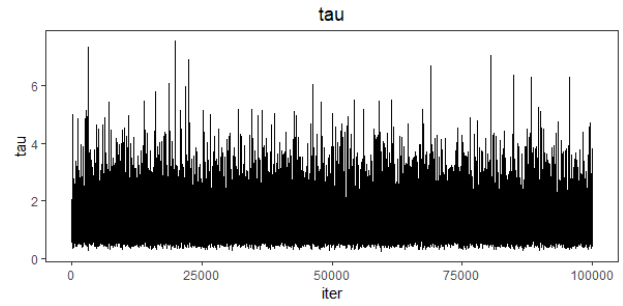
g. Sensitivity Analysis : Repeating Gibbs sampler for values of the hyperparameters (a, b, c, d)

1) half (a, b, c, d) = (1.5, 0.5, 30, 60)

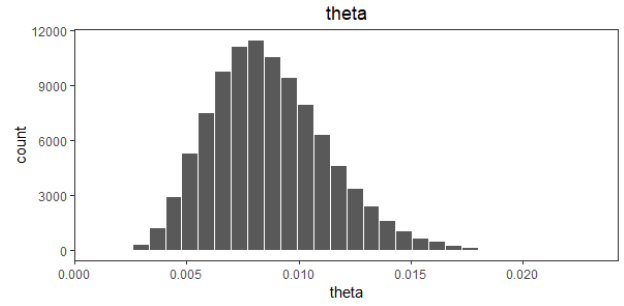
θ	τ

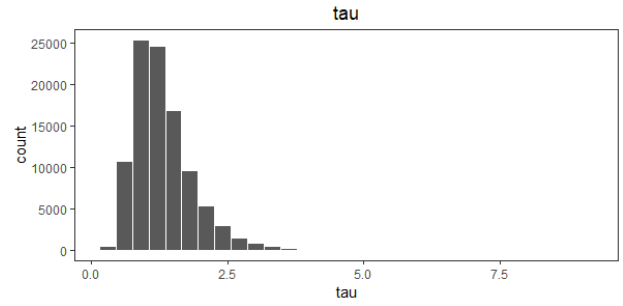
Sample paths



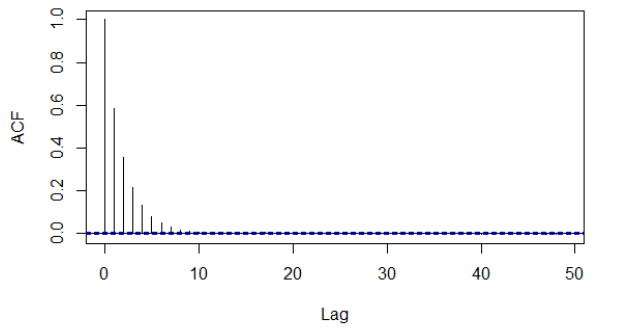


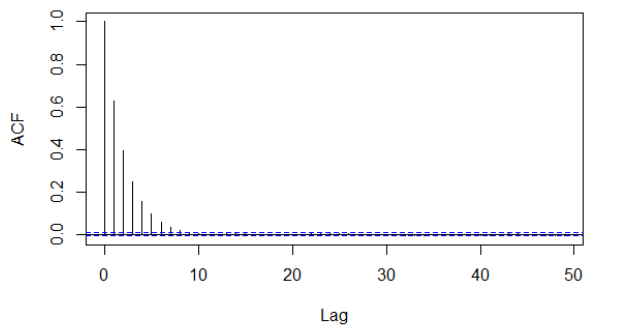
Histogram





Auto Correlation Function (ACF)



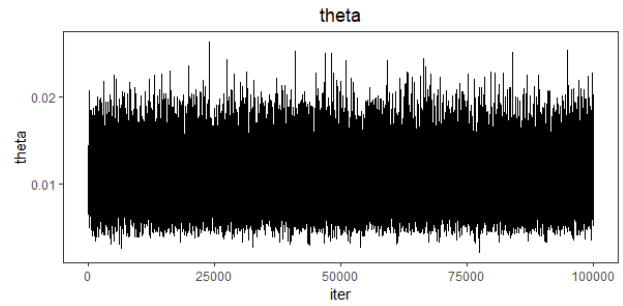


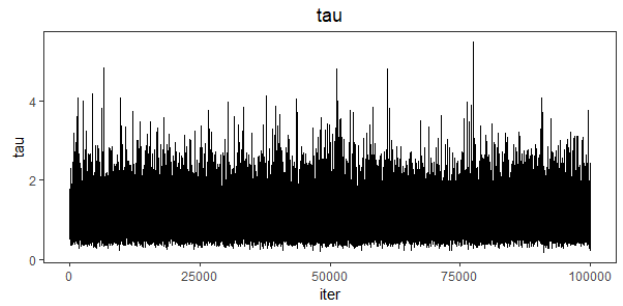
	mean	sd	min	Q1	Q2	Q3	max	95% Prob.Interval	
θ	0.00873	0.00263	0.00164	0.00685	0.00846	0.00873	0.01033	0.00435	0.01458
τ	1.30903	0.55348	0.24350	0.92470	1.20110	1.56930	7.54730	0.56456	2.68273

2) double (a, b, c, d) = (6, 2, 60, 120)

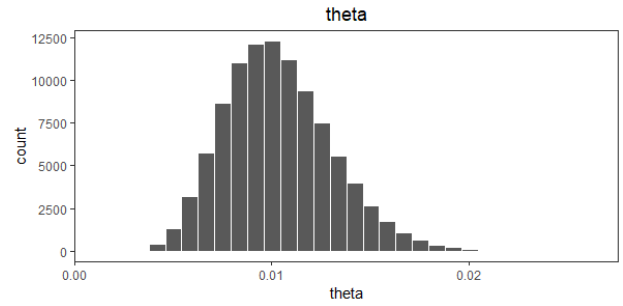
θ	τ

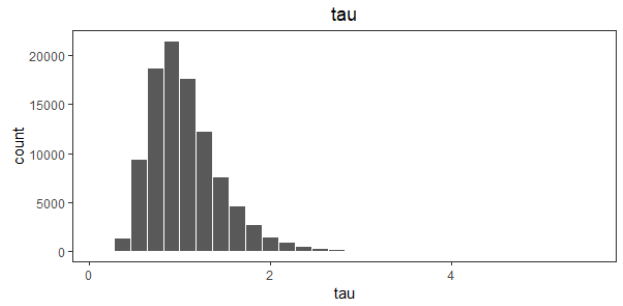
Sample paths



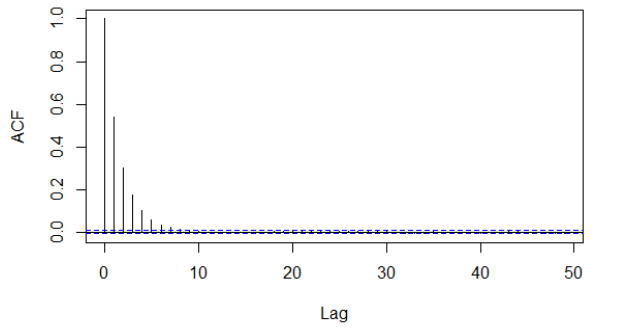


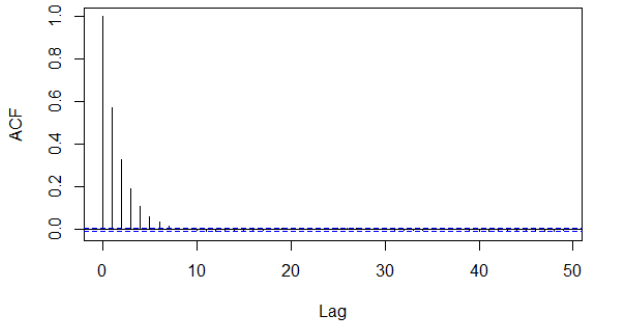
Histogram





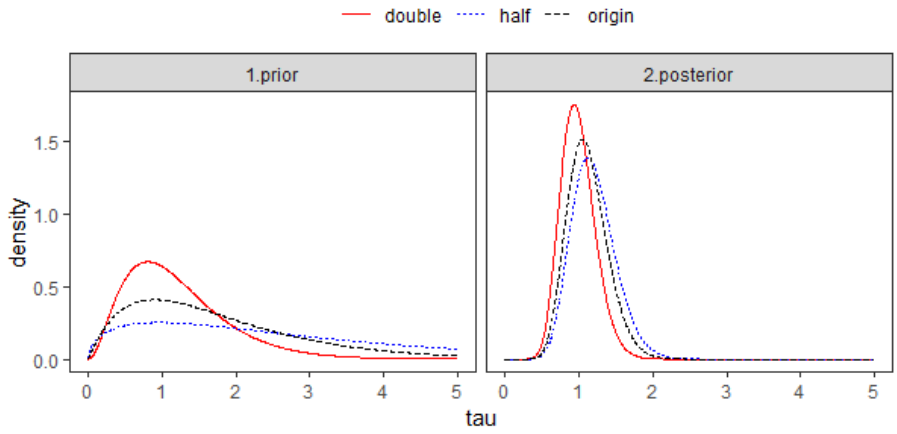
Auto Correlation Function (ACF)





	mean	sd	min	Q1	Q2	Q3	max	95% Prob.Interval	
θ	0.01034	0.00275	0.00208	0.00837	0.01010	0.01204	0.02633	0.00568	0.01638
τ	1.06032	0.39923	0.18830	0.77980	0.99000	1.26360	5.46930	0.49413	2.02485

hyperparameter를 0.5배, 2배로 각각 조정하였을 때 θ와 τ의 분포는 0.5배의 경우 조금 더 넓고 2배의 경우 좀 더 좁은 형태를 띄고 있지만 posterior 분포가 원래의 경우와 거의 달라지지 않았다.



Hyperparameter (a, b, c, d)의 값을 0.5배, 2배로 하여 sampling한 후 계산한 통계량의 결과를 보면 모든 경우 95% prediction interval이 1을 포함하여, 호르몬 치료의 효과가 존재한다고 할 수 없다는 검정 결론에는 변함이 없다. 위의 그래프는 각각의 parameter 값들에 따른 사전분포와 사후분포를 겹쳐 그린 것으로, hyperparameter에 따른 서로 다른 사전분포를 가정하더라도 관측치에 기반한 사후분포는 크게 달라지지 않는다. 따라서 사전분포에 대한 민감도는 크지 않다고 할 수 있다.

Ⅲ. Discussion

MCMC는 Markov Chain과 MC Simulation을 결합하여 관심 모수의 target 분포로부터 iid sample을 직접 draw할 수 없을 때 proposal distribution을 이용한 조건부 sampling을 통해 구성된 sample들이 서로 독립이 아니더라도 값들의 실제 분포와 같아지도록 하여 기댓값을 추정할 수 있는 뛰어난 방법이다. proposal distribution g의 선택에 따라, sample의 크기가 커짐에 따른 값들의 stationary distribution이 target distribution으로 수렴 여부가 달라진다. g는 target f와 유사한 형태여야 하며, f보다 두꺼운 꼬리를 가져야 적절한 proposal로서 sample을 draw할 수 있다.

Implementation 1에서 Independent chain의 Beta(1, 1), Random Walk chain의 U(1, -1)은 sample을 draw하기 위한 적절한 proposal distribution이 되었지만 Beta(2, 10)과 U(-0.01, 0.01)의 경우에는 그렇지 못하였다.

더 나아가 Implementation 2에서 Gibbs Sampling을 통해 모수가 여러 개인 결합 분포에서도 어려움 없이 조건부 sampling을 이용하여 joint distribution은 물론 각각 단일 모수의 marginal distribution 또한 target joint 또는 marginal distribution으로 수렴할 수 있음을 예제를 통해 알게 되었다.

[Appendix] R code

 # Example 7.2

```
data1 <- read.table("C:/Users/HG/Desktop/18-2/CS/Textbook/Datasets/mixture.dat", header = T) ; y <- data1$y
L <- function(d) { f = d*dnorm(y, 7, 0.5) + (1-d)*dnorm(y, 10, 0.5) ; return(prod(f)) }

n = 100000
# 1-1. independent chain
mychain1 <- function(a, b, n = 100000) {
  d = c(runif(1, 0, 1), numeric(n-1))
  for (i in 2:n) {
    d_ <- rbeta(1, a, b) ; R <- L(d_)/L(d[i-1])
    d[i] <- ifelse(rbinom(1, 1, min(1,R)) == 1, d_, d[i-1]) }
  return(d)
}

# 1) Prior: Beta(1, 1) = Unif(0, 1)
d1 <- mychain1(1, 1)
ggplot() + geom_line(aes(1:n, d1), color = "darkblue") + theme_test() + theme(plot.title = element_text(hjust = 0.5)) +
  labs(x = "iter", y = "delta", title = "Beta(1,1)")
ggplot() + theme_test() + theme(plot.title = element_text(hjust = 0.5)) +
  geom_histogram(aes(d1[-(1:500)]), fill = "darkblue", color = "white") + labs(x = "delta", title = "Beta(1,1)")

# 2) Prior: Beta(2, 10)
d2 <- mychain1(2, 10)
ggplot() + geom_line(aes(1:n, d2), color = "darkred") + theme_test() + theme(plot.title = element_text(hjust = 0.5)) +
  labs(x = "iter", y = "delta", title = "Beta(2,10)")
ggplot() + theme_test() + theme(plot.title = element_text(hjust = 0.5)) +
  geom_histogram(aes(d2[-(1:500)]), fill = "darkred", color = "white") +
  labs(x = "delta", title = "Beta(2,10)")

indep <- data.frame(d1, d2)[-(1:500),]

info1 <- data.frame(mean = apply(indep, 2, mean), sd = apply(indep, 2, sd),
                    lowerCI = apply(indep, 2, function(x) quantile(x, 0.025)),
                    upperCI = apply(indep, 2, function(x) quantile(x, 0.975)))
info1 %>% round(5)
summary(indep)

# 1-2. random walk chain
ilogit <- function(u) { exp(u)/(1+exp(u)) }

mychain2 <- function(b, n = 100000) {
  u = c(runif(1, 0, 1), numeric(n-1))
  for (i in 2:n) {
    u_ <- u[i-1] + runif(1, -b, b)
    R <- L(ilogit(u_))/L(ilogit(u[i-1])) * abs(ilogit(u_)/(1+exp(u_)))/abs(ilogit(u[i-1])/(1+exp(u[i-1])))
    u[i] <- ifelse(rbinom(1, 1, min(1,R)) == 1, u_, u[i-1]) }
  d <- ilogit(u) ; return(d)
}
d3 <- mychain2(1)
d4 <- mychain2(0.01)

ggplot() + geom_line(aes(1:n, d3), color = "darkblue") + theme_test() + theme(plot.title = element_text(hjust = 0.5)) +
  labs(x = "iter", y = "delta", title = "Unif(-1,1)") + lims(y = c(0.4, 0.9))
ggplot() + geom_line(aes(1:n, d4), color = "darkred") + theme_test() + theme(plot.title = element_text(hjust = 0.5)) +
  labs(x = "iter", y = "delta", title = "Unif(-0.01,0.01)") + lims(y = c(0.4, 0.9))

ggplot() + theme_test() + theme(plot.title = element_text(hjust = 0.5)) + labs(x = "delta", title = "Unif(-1,1)") + lims(x = c(0.5, 0.9)) +
  geom_histogram(aes(d3[-(1:500)]), fill = "darkblue", color = "white")
```

```
ggplot() + theme_test() + theme(plot.title = element_text(hjust = 0.5)) + labs(x = "delta", title = "Unif(-0.1,0.1)")  + lims(x = c(0.5, 0.9)) +
  geom_histogram(aes(d4[-(1:500)]), fill = "darkred", color = "white")

random <- data.frame(d3, d4)[-(1:500),]
info2 <- data.frame(mean = apply(random, 2, mean), sd = apply(random, 2, sd),
                    lowerCI = apply(random, 2, function(x) quantile(x, 0.025)),
                    upperCI = apply(random, 2, function(x) quantile(x, 0.975)))
info2 %>% round(5)
rbind(summary(d3[-(1:500)]), summary(d4[-(1:500)])) %>% round(5)
```

 # Problems 7.5

```
data2 <- read.table("C:/Users/HG/Desktop/18-2/CS/Textbook/Datasets/breastcancer.dat", header = T)
x <- data2$recurtime ; C <- data2$censored ; H <- data2$treatment

# a. summary & plot
summary(x[H==1&C==0])
summary(x[H==0&C==0])
summary(x[H==1&C==1])
summary(x[H==0&C==1])

a <- ggplot(data2) + theme_test() + scale_fill_brewer("treatment", palette = "Set1") + theme(plot.title = element_text(hjust = 0.5))
a + geom_density(aes(recurtime, group = treatment, fill = factor(treatment)), alpha = 0.5) + facet_wrap(~censored)

# b-c. sampling
gtheta <- function(tau, a=3, b=1, c=60, d=120) {
  rgamma(1, shape = a+sum((1-C)*(1-H))+sum((1-C)*H)+1, rate = c+sum((1-H)*x)+(d+sum(H*x))*tau) }
gtau <- function(theta, a=3, b=1, c=60, d=120) {
  rgamma(1, shape = b+sum((1-C)*H)+1, rate = (d+sum(H*x))*theta) }

mychain3 <- function(n = 10000, a=3, b=1, c=60, d=120) {
  theta <- c(gtheta(1, a, b, c, d), numeric(n-1)) ; tau <- c(gtau(theta[1], a, b, c, d), numeric(n-1))
  for (i in 2:n) { theta[i] <- gtheta(tau[i-1], a, b, c, d) ; tau[i] <- gtau(theta[i], a, b, c, d) }
  return(data.frame(theta, tau))
}

n = 100000 ; burnin = 1:500
gibbs1 <- mychain3(n) ; theta1 <- gibbs1$theta ; tau1 <- gibbs1$tau

plt <- ggplot() + theme_test() + theme(plot.title = element_text(hjust = 0.5)) + scale_fill_brewer("", palette = "Set1")
plt + geom_line(aes(1:n, theta1)) + labs(x = "iter", y = "theta", title = "theta")
plt + geom_line(aes(1:n, tau1)) + labs(x = "iter", y = "tau", title = "tau")

acf(theta1[-burnin], main="theta") ; acf(tau1[-burnin], main="tau")
plt + geom_histogram(aes(x = theta1[-burnin], y = ..density..), color = "white") + labs(x = "theta", title = "theta")
plt + geom_histogram(aes(x = tau1[-burnin], y = ..density..),  color = "white") + labs(x = "tau", title = "tau")

# d.summary statistics
gibbs1 <- gibbs1[-(1:500),] ; theta1 <- theta1[-(1:500)] ; tau1 <- tau1[-(1:500)]
summary(gibbs1)
summ1 <- data.frame(mean = apply(gibbs1, 2, mean), sd = apply(gibbs1, 2, sd),
                    lowerCI = apply(gibbs1, 2, function(x) sort(x)[length(x)*0.025]),
                    upperCI = apply(gibbs1, 2, function(x) sort(x)[length(x)*0.975]))
summ1 %>% round(5)

# e. tau plot
a=3 ; b=1 ; c=60 ; d=120

prior <- function(theta, tau, a=3, b=1, c=60, d=120) { theta^a * tau^b * exp(-c*theta-d*tau*theta) }
L <- function(theta, tau) {
  theta^(sum((1-C)*(1-H))+sum((1-C)*H)) * tau^(sum((1-C)*H)) * exp(-sum((1-H)*x)*theta-sum(H*x)*tau*theta) }
post <- function(theta, tau, a=3, b=1, c=60, d=120) { prior(theta, tau, a, b, c, d) * L(theta, tau) }
```

```
  tau <- seq(0, 6, length = 1000)
  prior_tau <- prior(mean(theta1), tau)   / integrate(prior, theta = mean(theta1), lower = 0, upper = Inf)$value
  post_tau <- post(mean(theta1), tau) / integrate(post, theta = mean(theta1), lower = 0, upper = Inf)$value

  plt + scale_color_brewer("", palette = "Set1", direction = -1) + labs(y = "density") +
    geom_area(aes(tau, prior_tau, fill = "prior"), color = "black", alpha = 0.3) +
    geom_area(aes(tau, post_tau, fill = "posterior"), color = "black", alpha = 0.3)

# g.
# g-1. half
  gibbs2 <- mychain3(n, a=3/2, b=1/2, c=60/2, d=120/2) ; theta2 <- gibbs2$theta ; tau2 <- gibbs2$tau

  plt + geom_line(aes(1:n, theta2)) + labs(x = "iter", y = "theta", title = "theta")
  plt + geom_line(aes(1:n, tau2)) + labs(x = "iter", y = "tau", title = "tau")
  acf(theta2[-burnin]) ; acf(tau2[-burnin])
  plt + geom_histogram(aes(theta2[-burnin]), color = "white") + labs(x = "theta", title = "theta")
  plt + geom_histogram(aes(tau2[-burnin]), color = "white") + labs(x = "tau", title = "tau")

  gibbs2 <- gibbs2[-burnin,] ; theta2 <- gibbs2$theta ; tau2 <- gibbs2$tau
  summary(gibbs2)
  summ2 <- data.frame(mean = apply(gibbs2, 2, mean), sd = apply(gibbs2, 2, sd),
                      lowerCI = apply(gibbs2, 2, function(x) quantile(x, 0.025)),
                      upperCI = apply(gibbs2, 2, function(x) quantile(x, 0.975)))
  summ2 %>% round(5)

# g-2. double
  gibbs3 <- mychain3(n, a=3*2, b=1*2, c=60*2, d=120*2) ; theta3 <- gibbs3$theta ; tau3 <- gibbs3$tau
  plt + geom_line(aes(1:n, theta3)) + labs(x = "iter", y = "theta", title = "theta")
  plt + geom_line(aes(1:n, tau3)) + labs(x = "iter", y = "tau", title = "tau")
  acf(theta3[-burnin]) ; acf(tau3[-burnin])
  plt + geom_histogram(aes(theta3[-burnin]), color = "white") + labs(x = "theta", title = "theta")
  plt + geom_histogram(aes(tau3[-burnin]), color = "white") + labs(x = "tau", title = "tau")

  gibbs3 <- gibbs3[-burnin,] ; theta3 <- gibbs3$theta ; tau3 <- gibbs3$tau
  summary(gibbs3)
  summ3 <- data.frame(mean = apply(gibbs3, 2, mean), sd = apply(gibbs3, 2, sd),
                      lowerCI = apply(gibbs3, 2, function(x) quantile(x, 0.025)),
                      upperCI = apply(gibbs3, 2, function(x) quantile(x, 0.975)))
  summ3 %>% round(5)

  tau <- seq(0, 5, length = 1000)
  prior_tau1 <- prior(mean(theta1), tau) / integrate(prior, theta = mean(theta1), lower = 0, upper = Inf)$value
  post_tau1 <- post(mean(theta1), tau) / integrate(post, theta = mean(theta1), lower = 0, upper = Inf)$value
  prior_tau2 <- prior(mean(theta2), tau, a=3/2, b=1/2, c=60/2, d=120/2) /
    integrate(prior, theta = mean(theta2), a=3/2, b=1/2, c=60/2, d=120/2, lower = 0, upper = Inf)$value
  post_tau2 <- post(mean(theta2), tau, a=3/2, b=1/2, c=60/2, d=120/2) /
    integrate(post, theta = mean(theta2), a=3/2, b=1/2, c=60/2, d=120/2, lower = 0, upper = Inf)$value
  prior_tau3 <- prior(mean(theta3), tau, a=3*2, b=1*2, c=60*2, d=120*2) /
    integrate(prior, theta = mean(theta3), a=3*2, b=1*2, c=60*2, d=120*2, lower = 0, upper = Inf)$value
  post_tau3 <- post(mean(theta3), tau, a=3*2, b=1*2, c=60*2, d=120*2) /
    integrate(post, theta = mean(theta3), a=3*2, b=1*2, c=60*2, d=120*2, lower = 0, upper = Inf)$value

  compare <- data.frame(par = c(rep("origin", length(tau)), rep("half", length(tau)), rep("double", length(tau))),
                        dist = c(rep("1.prior", 3*length(tau)), rep("2.posterior", 3*length(tau))),
                        density = c(prior_tau1, prior_tau2, prior_tau3, post_tau1, post_tau2, post_tau3))

  ggplot(compare) + theme_test() +
    scale_linetype_discrete("") + scale_color_manual("", values = c(origin = "black", half = "blue", double = "red")) +
    geom_line(aes(rep(tau, 6), density, color = par, linetype = par)) + facet_wrap(~dist) + labs(x = "tau") + theme(legend.position = "top")
```