MCMC Class

A Long journey to Markov Chain Monte-Carlo
Bayesian Estimation with Metropolis Rule.

2021/09

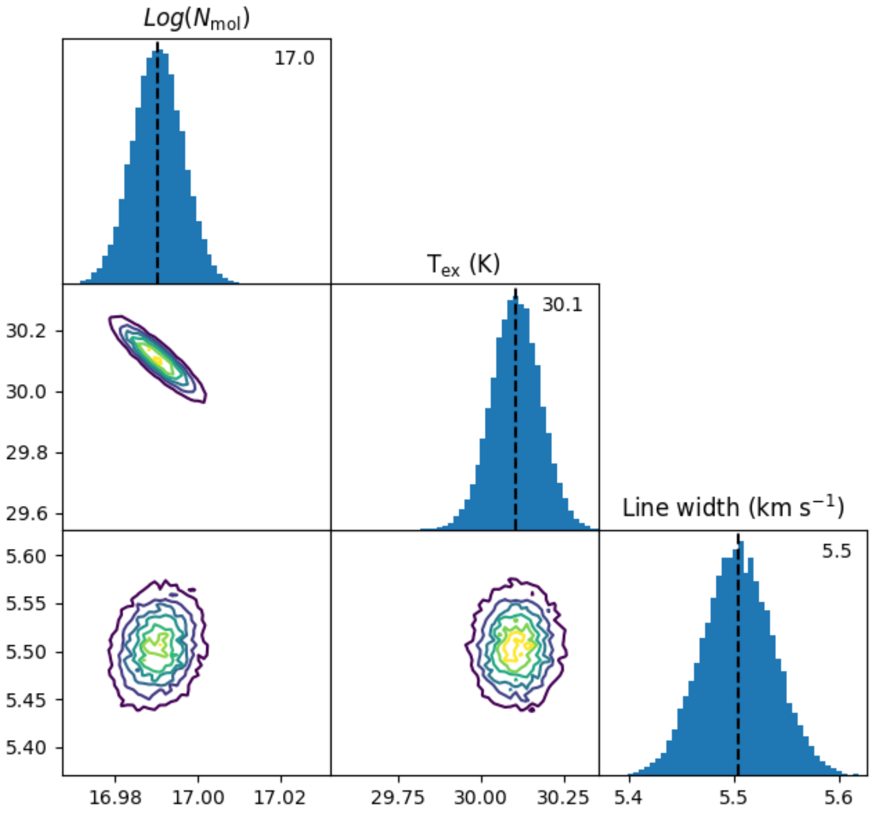
Hyeong-sik Yun
Star formation lab., Kyung Hee University

Objective

- Markov Chain Monte-Carlo (MCMC) 방법을 이해하기위해 알아야하는 기본적인 통계 지식에 대해 간략하게 익힌다.
- MCMC 방법을 이용한 data fitting이 어떤 방식으로 이루어지는지 이해한다.
- MCMC 방법을 이용한 data fitting을 주어진 test problem을 바탕으로 사용방법과 결과 해석을 시도한다.

❖스스로 MCMC code를 build 해보고 사용하는 것을 최종 목표로 한다.

What is MCMC?



- One of the method to find the best-fit model?
- What is the output of the MCMC method?
- What is the MCMC?

Contents

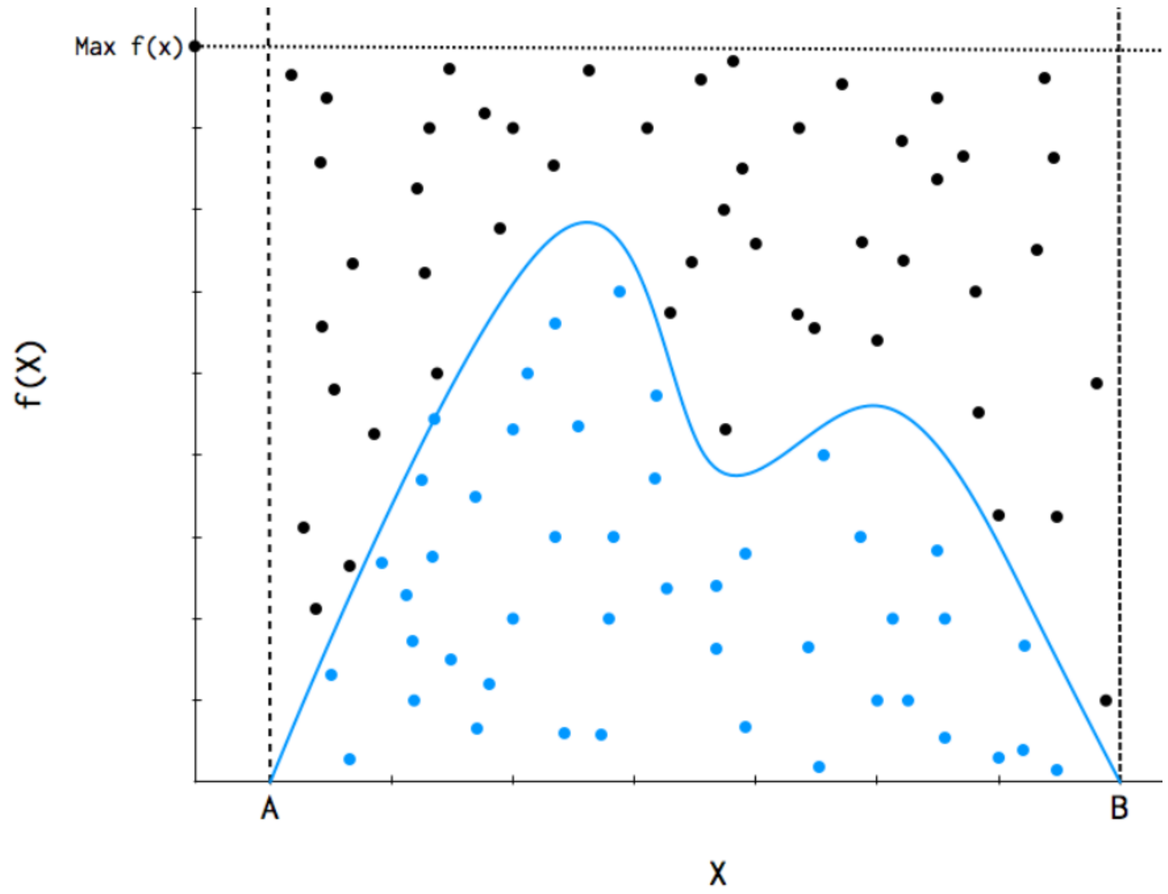
Monte-Carlo

- 통계적인 수치를 얻기 위한 simulation.
- 무한한 시도를 거쳐야 답을 얻을 수 있는 문제에 대해 유한한 시도로 정답을 추정하고자 하는 것.
- 통계적인 특성을 이용해 무수한 시도를 하는 분석을 통틀어 지칭함.
- Ex) Monte-Carlo sampling, Bootstrap method, …

❖Monte-Carlo, Monaco: 지중해 연안에 위치한 카지노와 도박장으로 유명한 도시.

Monte-Carlo Sampling

- Sampling: 모집단으로부터 조건에 맞게 표본을 추출하는 것.



Markov chain

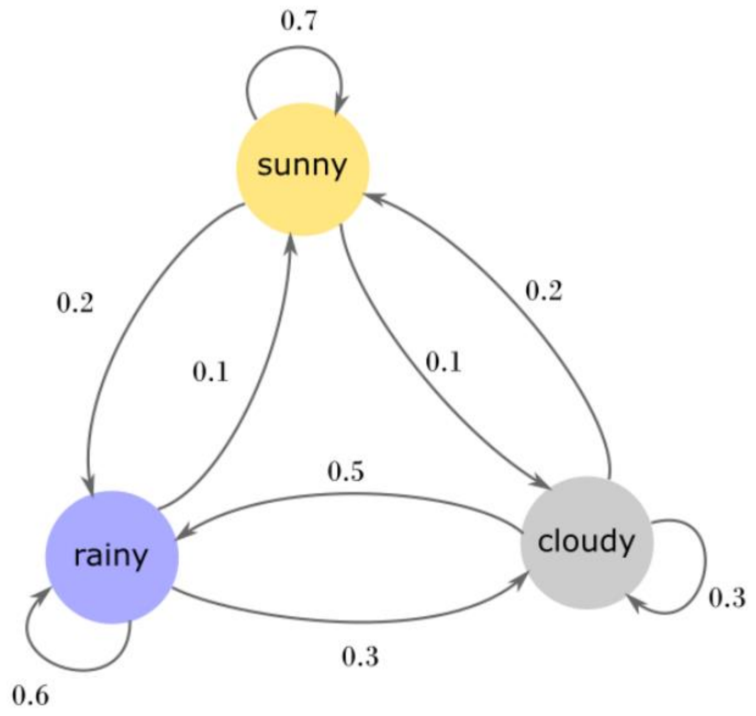
- 과정을 설명하는 통계적 model에서 많이 사용되는 방법
- 연속되는 state 를 기술하는 확률 모델이며, 각 state 의 확률은 직전 state 에만 영향을 받는다.

Ex1) 어제 순대국밥을 먹었으니
오늘 점심은 다른걸 먹자.

직전의 state ->

다음 state의 확률에 영향을 줌

Ex2)



Markov Chain Monte-Carlo (MCMC)

- 무한한 시도를 거쳐야 답을 얻을 수 있는 문제에 대해 유한한 시도로 정답을 추정하고자 하는 것 (**Monte-Carlo**).
- 일반 **Monte-Carlo** 방법에 비해 좀더 적은 시도로 정답을 추정하고자 **Markov Chain**을 활용하여 접근하는 방법.

특징

1. 첫 sample(θ_0)로부터 다음 sample(θ_1)이 propose 된다.
2. Propose된 sample(θ_i)은 **특정 기준에 따라 Accept 또는 Reject**한다.
3. Sample(θ_i)이 Accept 된 경우, **다음 sample(θ_{i+1})은 직전 sample(θ_i)에서 propose** 된다 (chain).
4. **2-3 과정을 무수히 많이 반복한다.**

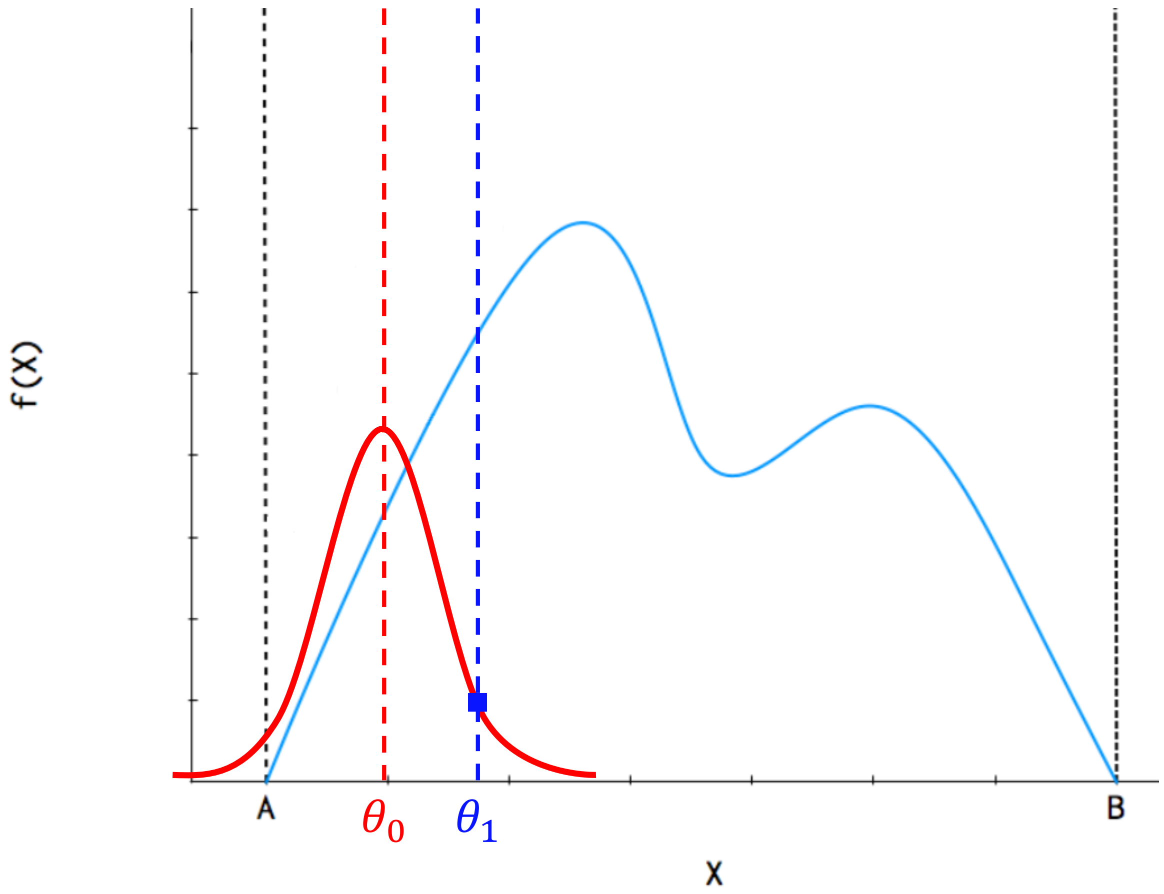
Markov Chain Monte-Carlo (MCMC)

- MCMC는 fitting method의 이름이 아니다.
- Output of MCMC = **Posterior** (Accept 된 sample의 집합)

- 수집된 Posterior 를 이용하여 **알아내고자 하는 정답을 추정하는 것** (데이터를 잘 설명하는 분포 -> **fitting**; 뒤에서 계속)

The MCMC sampling

1. Random initial sample(θ_0)
2. Generate a sample(θ_1) from θ_0
   - Normal probability density function (PDF) is generally used.
   - Propose θ_1
3. Accept or Reject the propose with **a given rule(sampler)**.

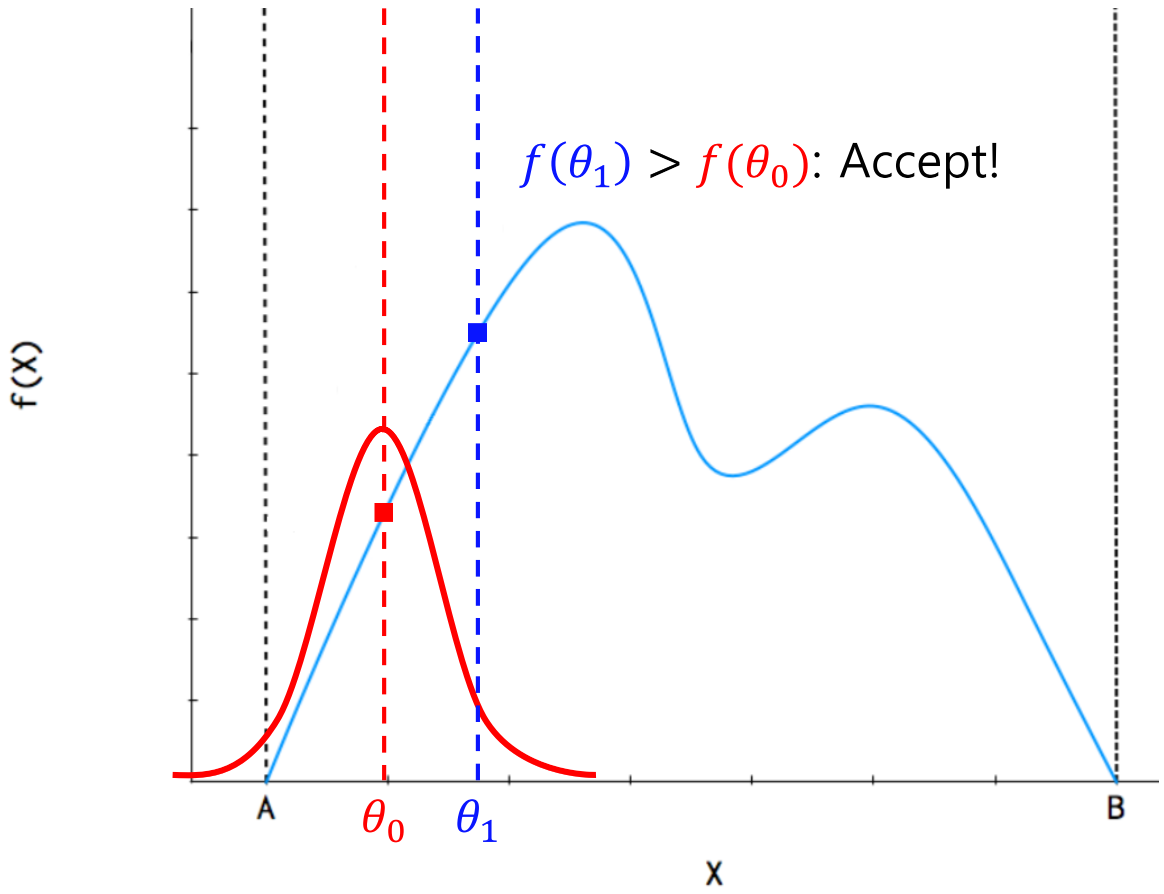


❖the simplest and commonly used sampler: **Metropolis rule**

The MCMC sampling with Metropolis rule

1. Random initial sample(θ_0)
2. Generate a sample(θ_1) from θ_0
   - Normal probability density function (PDF) is generally used.
   - Propose θ_1
3. Judge the acceptance

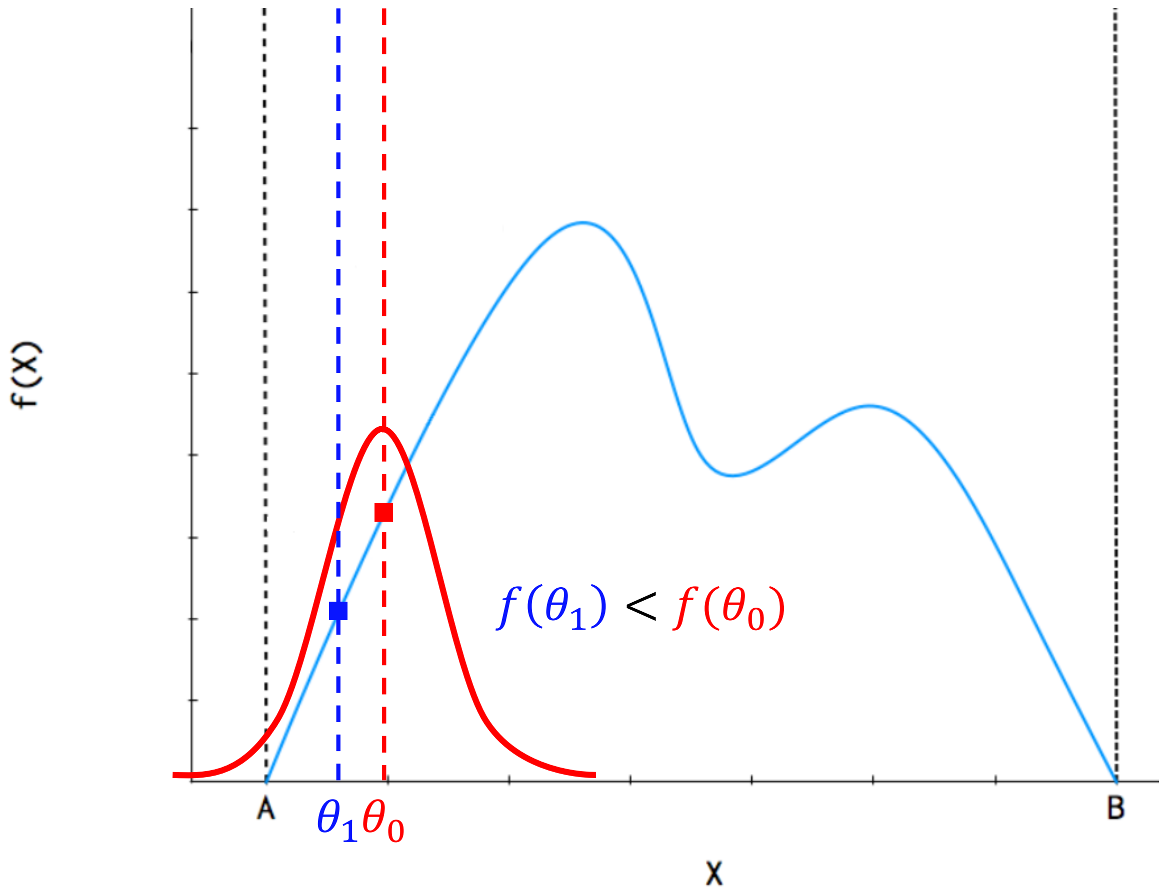
$$\begin{cases} f(\theta_1) \geq f(\theta_0): Accept \\ \end{cases}$$



The MCMC sampling with Metropolis rule

1. Random initial sample(θ_0)
2. Generate a sample(θ_1) from θ_0
   - Normal probability density function (PDF) is generally used.
   - Propose θ_1
3. Judge the acceptance

$$\begin{cases} f(\theta_1) \geq f(\theta_0): Accept \\ f(\theta_1) < f(\theta_0): Accept\ with\ p = \dfrac{f(\theta_1)}{f(\theta_0)} \end{cases}$$



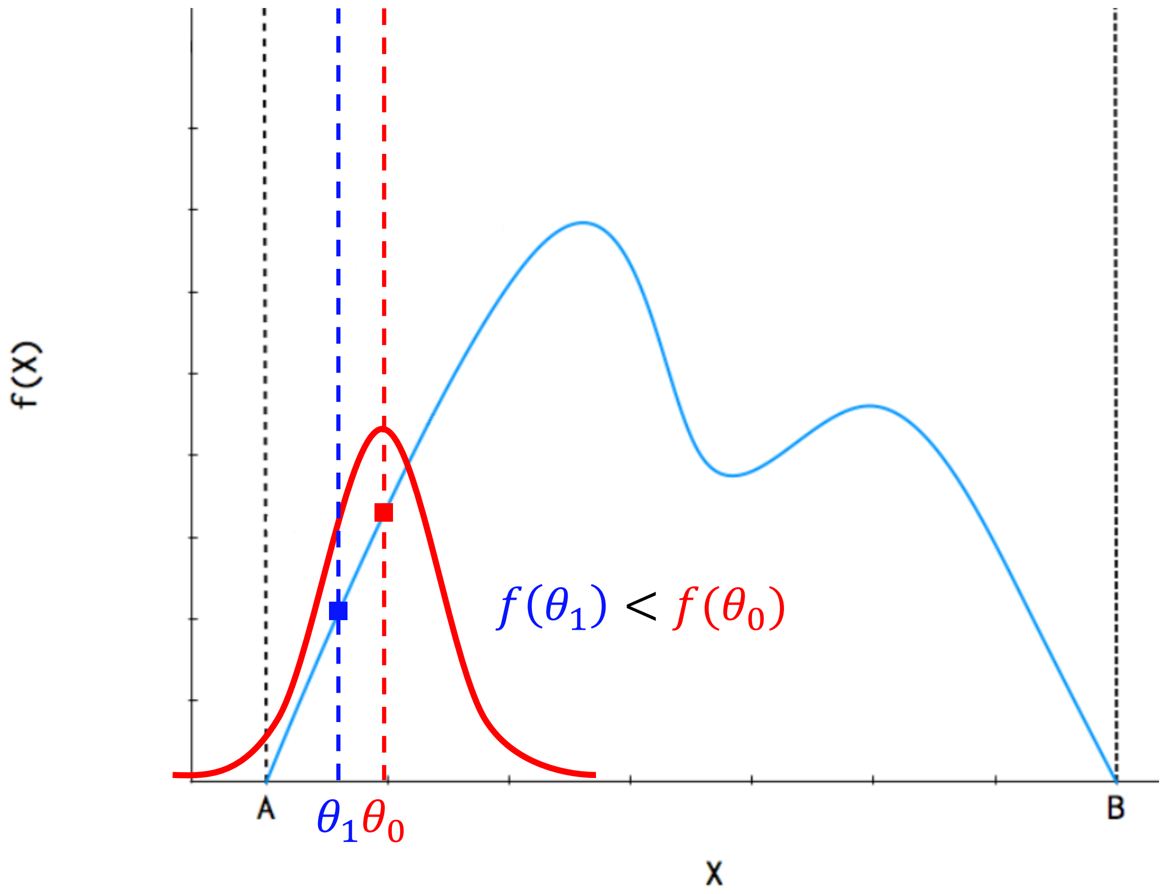
The MCMC sampling with Metropolis rule

- Judge the acceptance

$$\begin{cases} f(\theta_1) \geq f(\theta_0): Accept \\ f(\theta_1) < f(\theta_0): Accept\ with\ p = \dfrac{f(\theta_1)}{f(\theta_0)} \end{cases}$$

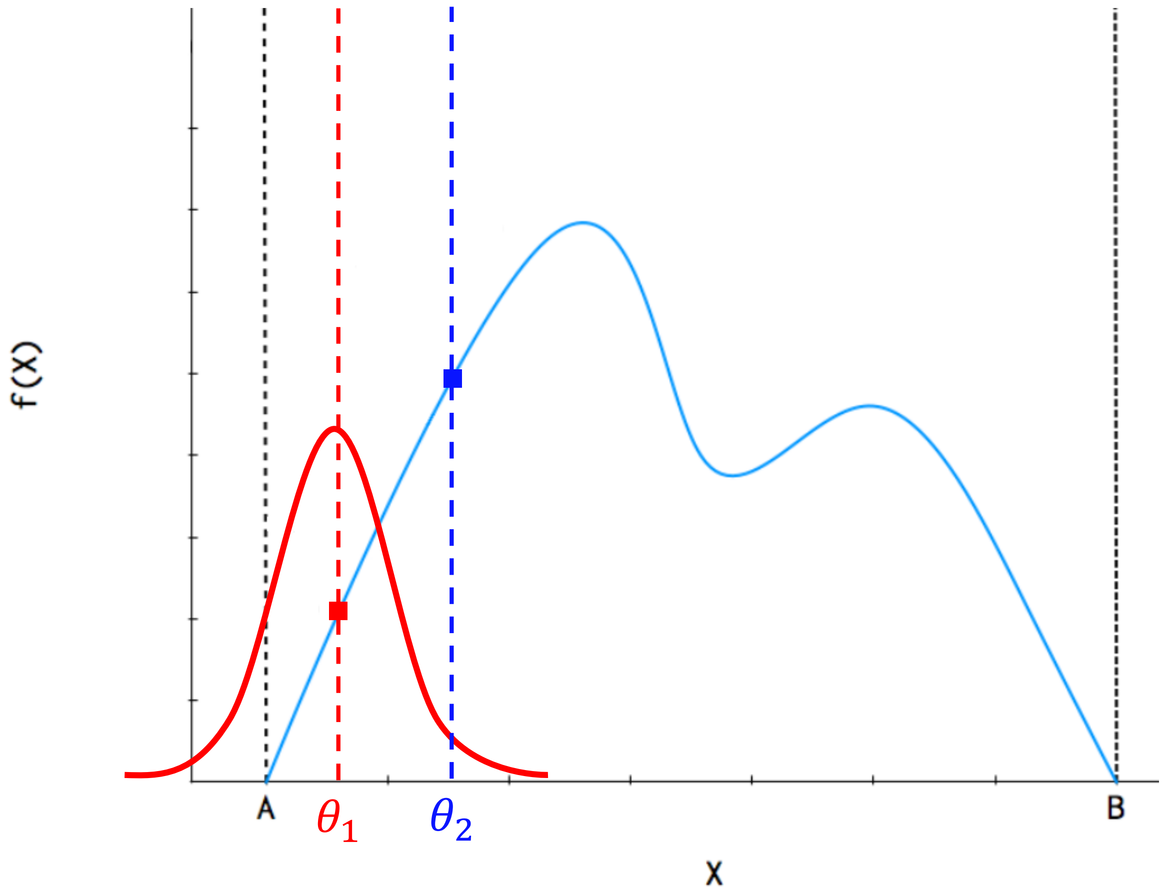
- Why we accept the propose stochastically?
  1. Allow to escape the local minima
  2. Provide a good chance to find the deep minima

Ex) Markov Chain Monte Carlo - 공돌이의 수학정리노트 (angeloyeo.github.io)



The MCMC sampling with Metropolis rule

1. Random initial sample(θ_0)
2. Generate a sample(θ_1) from θ_0
   - Normal probability density function (PDF) is generally used.
   - Propose θ_1
3. Judge the acceptance
4. Generate new sample (θ_{i+1}) from the accepted sample (posterior; θ_i)
5. Repeat 3-4



Likelihood

- Definition:

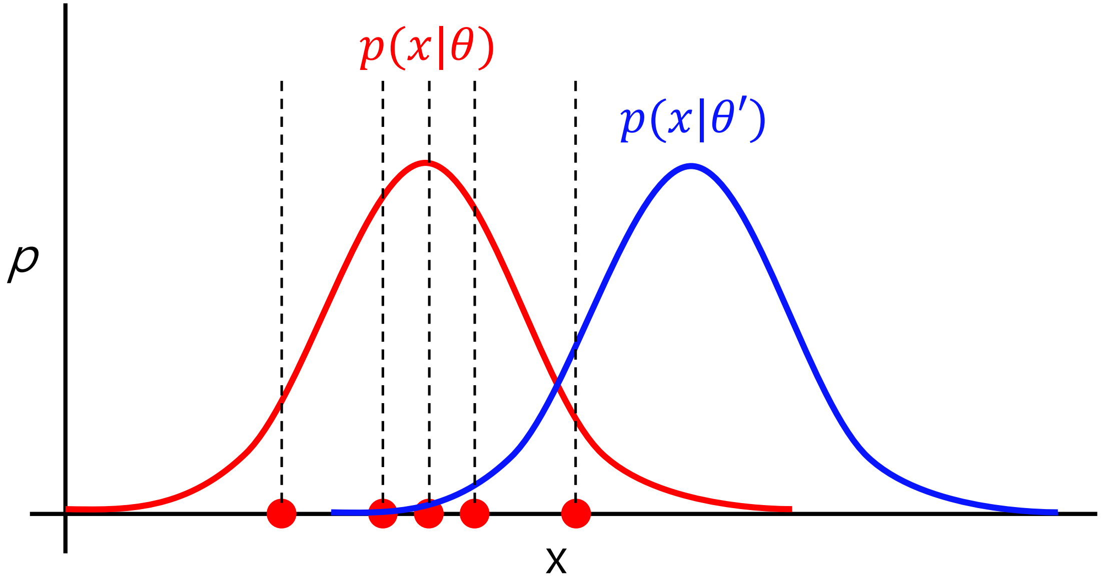
$$P(x|\theta) = \prod_{k=1}^{n} p(x_k|\theta) = p(x_1|\theta) * p(x_2|\theta) * \cdots * p(x_n|\theta)$$

  주어진 parameter(θ)로 정의된 확률밀도함수, $p(x|\theta)$가 있을 때, n개의 sample(x_k)에 대한 확률밀도의 곱

- 보통은 계산의 편의성 때문에 **Log-likelihood를 주로 사용한다**.

$$L(x|\theta) = \log P(x|\theta) = \log\left(\prod_{k=1}^{n} p(x_k|\theta)\right) = \sum_{k=1}^{n} \log p(x_k|\theta)$$

Maximum Likelihood Estimation (MLE)



❖알지 못하는 정규분포인 모 분포로부터 5개의 sample을 얻어냈다. 이 sample이 얻어진 모분포를 추정하기 위하여 Likelihood를 이용한다.

$$P(x|\theta) > P(x|\theta') \quad \& \quad L(x|\theta) > L(x|\theta')$$

- Likelihood가 Maximum이 되는 θ를 찾아 모 분포의 parameter를 추정하는 방법.

Bayes' theorem (베이즈 정리)

- Definition:

$$p(H|E) = \frac{p(E|H)\,p(H)}{p(E)}$$

[p: probability, H: hypothesis (the event interested in), E: evidence (new information)]

❖베이즈 정리에서 확률은 **'어떤 state가 가지는 신뢰도'**로 생각하는 것이 이해에 도움이 된다.

Bayes' theorem (베이즈 정리)

- Definition:

$$p(H|E) = \frac{p(E|H)\,p(H)}{p(E)}$$

[p: probability, H: hypothesis (the event interested in), E: evidence (new information)]

- 불확실성을 내포하는 수치를 기반으로 한 사전확률이 있을 때, 이에대한 **추가 사건을 관찰한 이후에 사전확률의 신뢰도를 갱신하는 방법**
  - $p(H)$: 기존에 알려져 있던 확률 (사전 확률; prior)
  - $p(H|E)$: 새로운 정보인 evidence를 얻은 후 확률 (사후 확률; posterior)

Bayes' theorem (베이즈 정리)

- Example problem1 ([베이즈 정리의 의미 - 공돌이의 수학정리노트 (angeloyeo.github.io)](angeloyeo.github.io))

  질병 A의 발병률은 0.1%로 알려져 있다. 이 질병에 실제로 감염된 환자에게 검사를 실시 하였을 때 양성이 나올 확률은 99% 이고, 질병에 감염되지 않은 사람에게 검사를 실시 하였을 때 음성이 나올 확률은 98% 이다.

  만약 어떤 X라는 사람이 검진에서 양성이 나왔을 때, **이 사람이 정말로 질병에 걸렸을 확률은?**

  이 문제에서, Hypothesis와 Evidence는 무엇인가?

Bayes' theorem (베이즈 정리)

- Example problem1 ([베이즈 정리의 의미 - 공돌이의 수학정리노트 (angeloyeo.github.io)](angeloyeo.github.io))

  질병 A의 발병률은 0.1%로 알려져 있다. 이 질병에 실제로 감염된 환자에게 검사를 실시 하였을 때 양성이 나올 확률은 99% 이고, 질병에 감염되지 않은 사람에게 검사를 실시 하였을 때 음성이 나올 확률은 98% 이다.

  만약 어떤 X라는 사람이 검진에서 양성이 나왔을 때, **이 사람이 정말로 질병에 걸렸을 확률은?**

이 문제에서, Hypothesis와 Evidence는 무엇인가?
- H: X는 질병 A의 **감염자**다
- E: X는 질병 A에 대해 **양성판정**을 받았다.

$$p(H|E) = \frac{p(E|H)p(H)}{p(E)} = \frac{p(E|H)p(H)}{p(E|H)p(H) + p(E|H^c)p(H^c)}$$

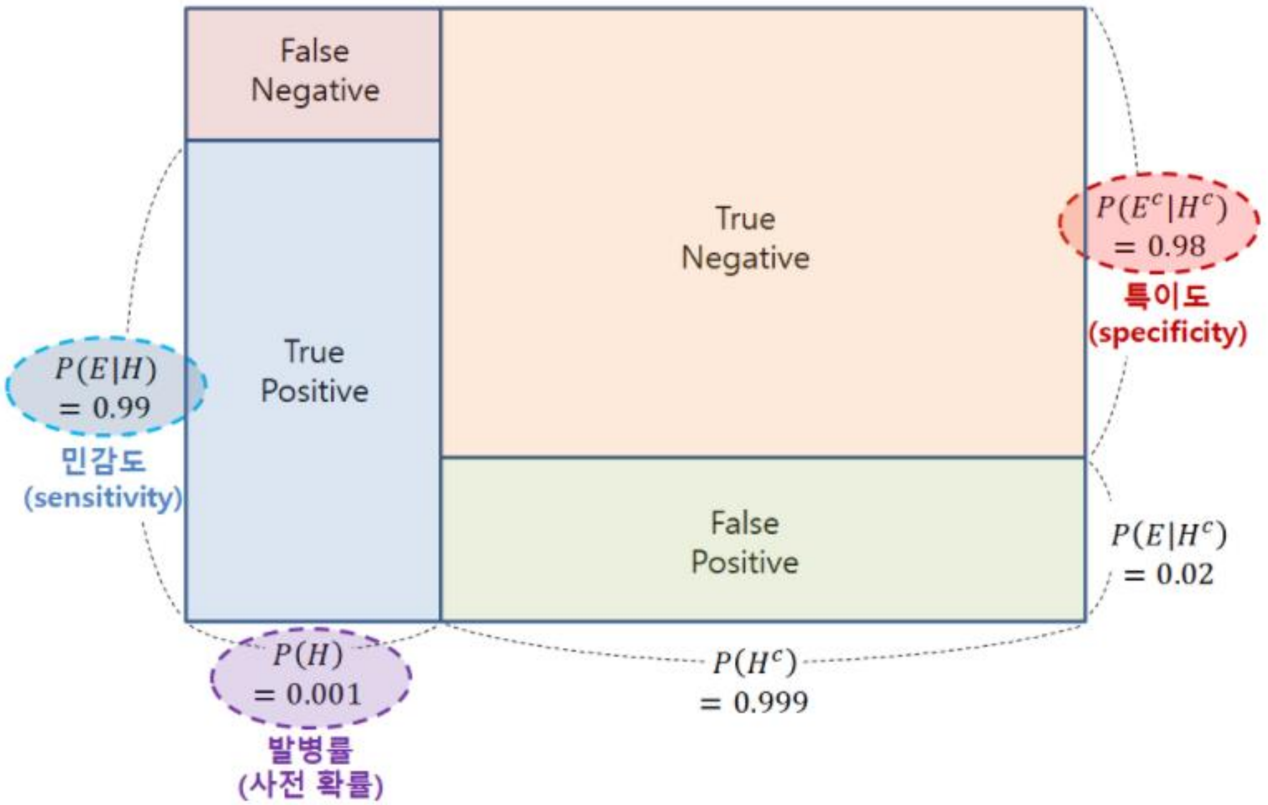
Bayes' theorem (베이즈 정리)

- Example problem1 ([베이즈 정리의 의미 - 공돌이의 수학정리노트 (angeloyeo.github.io)](angeloyeo.github.io))

$$p(H|E) = \frac{p(E|H)p(H)}{p(E)} = \frac{p(E|H)p(H)}{p(E|H)p(H) + p(E|H^c)p(H^c)}$$

$p(H|E)$: **양성 판정**을 받았을 때 **감염자**일 확률 (신뢰도)
$p(E|H)$: **감염자**일 때 **양성판정**을 받을 확률
$p(H)$: **감염자**일 확률
$p(E)$: **양성판정**을 받을 확률

$$p(H|E) = 0.047 = 4.7\%$$



Bayes' theorem (베이즈 정리)

- Example problem2 ([베이즈 정리의 의미 - 공돌이의 수학정리노트 (angeloyeo.github.io)](angeloyeo.github.io))

  만약 어떤 X라는 사람이 검진에서 양성이 나왔을 때, 이 사람이 정말로 질병에 걸렸을 확률은 4.7%이다. 그렇다면 한번 더 검진하여 양성이 나왔을 때, X가 질병 A의 감염자일 확률은?

Bayes' theorem (베이즈 정리)

- Example problem2 ([베이즈 정리의 의미 - 공돌이의 수학정리노트 (angeloyeo.github.io)])

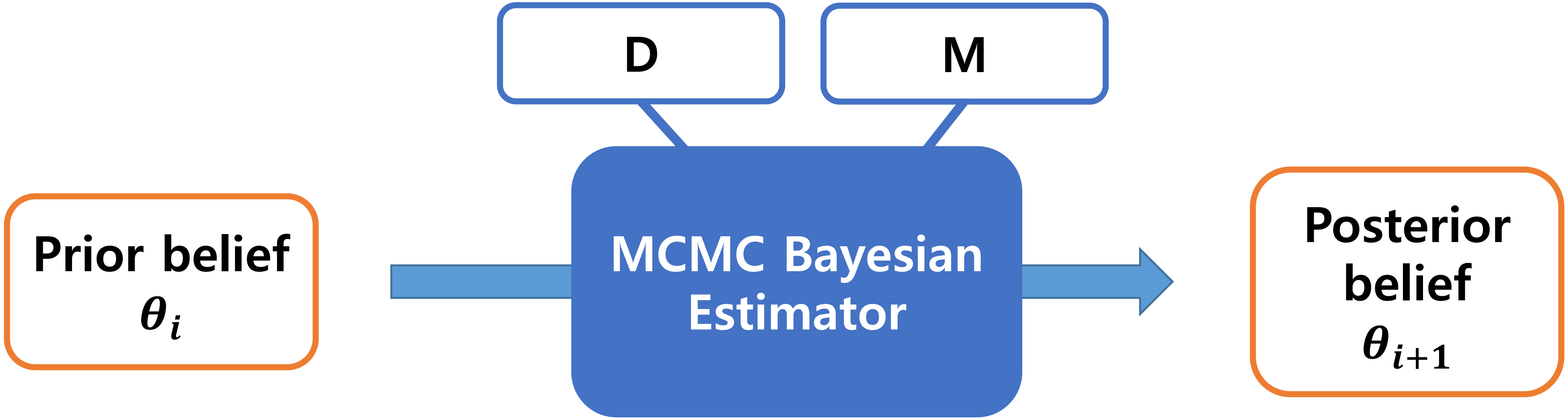
  만약 어떤 X라는 사람이 검진에서 양성이 나왔을 때, 이 사람이 정말로 질병에 걸렸을 확률은 4.7%이다. 그렇다면 한번 더 검진하여 양성이 나왔을 때, X가 질병 A의 감염자일 확률은?

$$p(H|E) = \frac{p(E|H)p(H)}{p(E)} = \frac{p(E|H)p(H)}{p(E|H)p(H) + p(E|H^c)p(H^c)}$$

- E: X는 두번째 검진에서 양성이 나왔다.
- H: X는 질병 A의 감염자다.
- *Prior* : 0.047
- *Posterior* : $p(H|E) = 0.709$

MCMC Bayesian Estimation with Metropolis rule

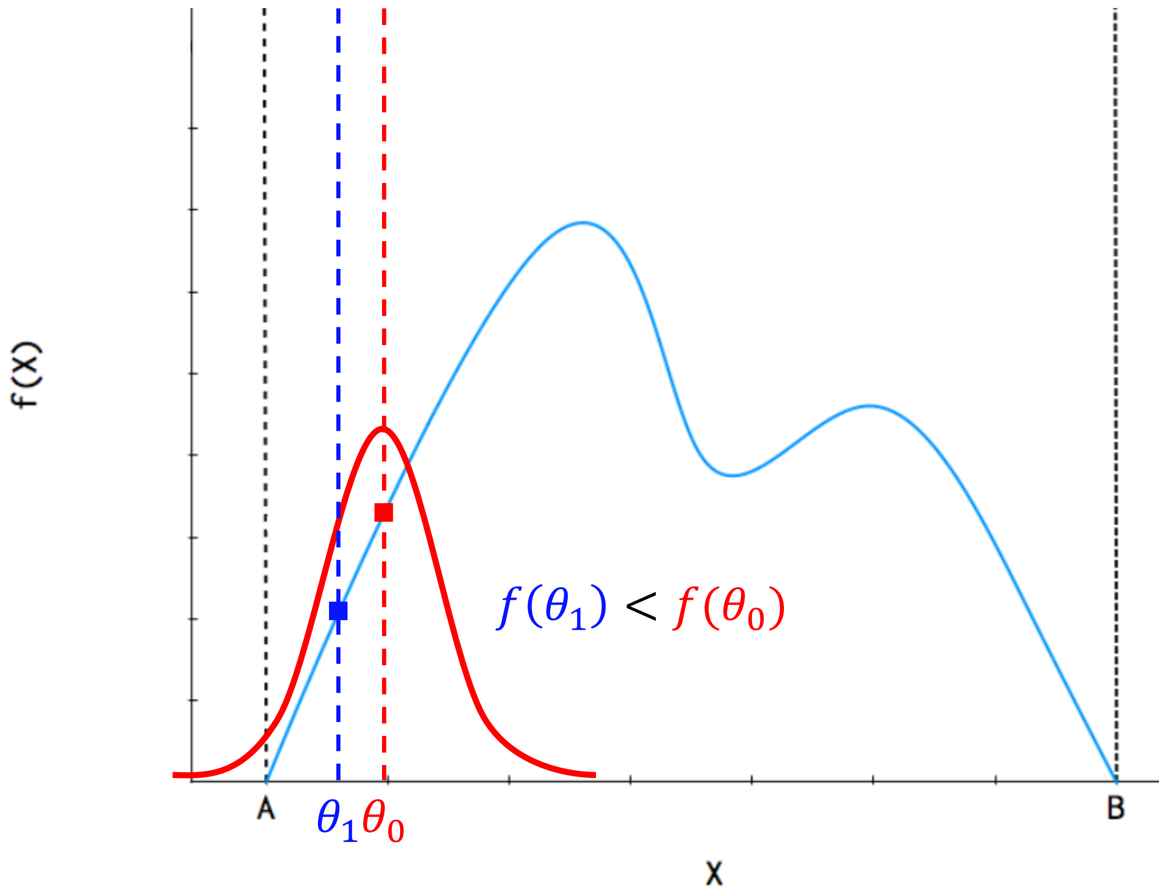
- θ를 parameter로 한 model(M)로 주어진 data(D)를 설명할 수 있을 때, MCMC를 이용하여 data를 가장 잘 설명할 수 있는 θ를 추정하는 방법.



The MCMC sampling with Metropolis rule

1. Random initial sample(θ_0)
2. Generate a sample(θ_1) from θ_0
   - Normal probability density function (PDF) is generally used.
   - Propose θ_1
3. Judge the acceptance

$$\begin{cases} f(\theta_1) \geq f(\theta_0): Accept \\ f(\theta_1) < f(\theta_0): Accept\ with\ p = \dfrac{f(\theta_1)}{f(\theta_0)} \end{cases}$$



MCMC Bayesian Estimation with Metropolis rule

1. Define initial parameters, θ_0.
2. Generate and propose a small jump $(\theta_0 + \Delta\theta)$
3. Judge the acceptance with metropolis rule

$$\begin{cases} f(\theta_1) \geq f(\theta_0):\ Accept \\ f(\theta_1) < f(\theta_0): Accept\ with\ p = \dfrac{f(\theta_1)}{f(\theta_0)} \end{cases}$$

❖어떤 함수가 $f(\theta)$에 쓰이는가?

'θ가 데이터를 잘 설명하는지'를 판단할 수 있는 지표 (신뢰도)

Bayes' theorem (베이즈 정리)

- Definition:

$$p(H|E) = \frac{p(E|H)\,p(H)}{p(E)}$$

[p: probability, H: hypothesis (the event interested in), E: evidence (new information)]

- 각 항목의 의미
  - Prior/posterior: Hypothesis가 맞을 확률(신뢰도)
  - Evidence: 새로운 정보 (Evidence의 신뢰도)
  - $p(E|H)$: Hypothesis가 맞을 때 Evidence가 맞을 확률 (신뢰도)

MCMC Bayesian Estimation with Metropolis rule

posterior　　Likelihood　　prior

$$p(\theta|D,M) = \frac{p(D|\theta,M)p(\theta|M)}{p(D|M)} \text{ (Bayes' theorem)}$$

Evidence Horrible integration over θ;
Independent from θ

- 각 항목의 의미
  - Prior/posterior: 주어진 파라미터(θ)를 이용해 만들어진 Model이 관측 Data를 잘 설명할 확률 (신뢰도)
  - Evidence: 임의의 Residual(Data – Model)이 Noise distribution을 따를 확률 (신뢰도)
  - Likelihood: 주어진 파라미터(θ)를 이용해 Residual을 구했을 때, 이 Residual이 noise distribution을 따를 확률 (신뢰도). 주로 normal white noise를 가정.

❖언제나 normal white noise가 사용되는 것은 아니다. Noise의 알려진 확률분포 (pdf)에 따라 다른 pdf를 사용해야 한다.

MCMC Bayesian Estimation with Metropolis rule

prior

$$p(\theta|D,M) = \frac{p(D|\theta,M)p(\theta|M)}{p(D|M)} \text{ (Bayes' theorem)}$$

Evidence

- MCMC Bayesian Estimation에서는 **Evidence**는 잘 고려하지 않는다.
- 주로 Metropolis rule을 적용한 경우, **Evidence는 상쇄된다.**

$$\frac{f(\theta_{i+1})}{f(\theta_i)} = \frac{p(\theta_{i+1}|D,M)}{p(\theta_i|D,M)} = \frac{p(D|\theta_{i+1},M)p(\theta_{i+1}|M)}{p(D|\theta_i,M)p(\theta_i|M)}$$

❖Evidence 는 θ에 대해 상수이기 때문에

$$p(\theta_i|D,M) \propto p(D|\theta_i,M)p(\theta_i|D)$$

❖여러 연구에서는 **Prior**를 고려하지 않음; $p(\theta_i|D,M) \propto p(D|\theta_i,M)$ **= Likelihood**

Likelihood

- Definition:

$$P(x|\theta) = \prod_{k=1}^{n} p(x_k|\theta) = p(x_1|\theta) * p(x_2|\theta) * \cdots * p(x_n|\theta)$$

- MCMC에서는 **Residual(r_i)**이 **Noise distribution**과 잘 일치하는지를 판단.
- 대부분의 천문학 연구에서는 Error를 **normal white noise를 가정**한다.

-> Gaussian likelihood function을 많이 사용.

$$P(D|\theta_i, M) = \prod_{k=1}^{n} \frac{1}{(2\pi\sqrt{\sigma_k})^2} exp\left(-\frac{1}{2}\frac{r_k^2}{\sigma_k^2}\right); r_k = D_k - M(\theta_i)_k$$

MCMC Bayesian Estimation with Metropolis rule

Likelihood

$$p(\theta|D,M) \propto p(D|\theta,M)p(D|\theta,M) = P(D|\theta,M)p(D|\theta,M)$$

Observed data

$$P(D|\theta_i,M) = \prod_{k=1}^{n} \frac{1}{(2\pi\sqrt{\sigma_k})^2} exp\left(-\frac{1}{2}\frac{r_k^2}{\sigma_k^2}\right); r_k = D_k - M(\theta_i)_k$$

Synthetised data

Measured uncertainty

❖ 종종 편의를 위해 Log-likelihood를 사용

$$p(\theta|D,M) \propto L(D|\theta,M)p(D|\theta,M)$$

$$L(D|\theta_i,M) = \sum_{k=1}^{n} \log\left(\frac{1}{(2\pi\sqrt{\sigma_k})^2} exp\left(-\frac{1}{2}\frac{r_k^2}{\sigma_k^2}\right)\right) = \frac{1}{2}\sum_{k=1}^{n}\left(-\log 2\pi - \log\sigma_k^2 - \frac{r_k^2}{\sigma_k^2}\right)$$

❖만약 σ_k가 일정(constant uncertainty)하다면,

$$L(D|\theta_i,M) = c - {}^{1}/_{2} * \mathrm{X}^2$$

Chi-square (X^2)

MCMC Bayesian Estimation with Metropolis rule

1. Define initial parameters, θ_0.
2. Generate and propose a small jump $(\theta_0 + \Delta\theta)$
3. **Judge the acceptance with metropolis rule**
4. Generate and propose next small jump $(\theta_{i+1} + \Delta\theta)$ based on the previous step $(\theta_i + \Delta\theta)$.
5. Repeat 3rd-4th steps

General case

If $p(\theta_{i+1}|D,M) \geq p(\theta_i|D,M)$, accept the jump

If $p(\theta_{i+1}|D,M) < p(\theta_i|D,M)$, accept the jump with

$$p = \frac{p(\theta_{i+1}|D,M)}{p(\theta_i|D,M)}$$

Ignoring priors

If $L(\theta_{i+1}|D,M) \geq L(\theta_i|D,M)$, accept the jump

If $L(\theta_{i+1}|D,M) < L(\theta_i|D,M)$, accept the jump with

$$p = \frac{L(\theta_{i+1}|D,M)}{L(\theta_i|D,M)}$$

Ignoring priors and assuming normal errors

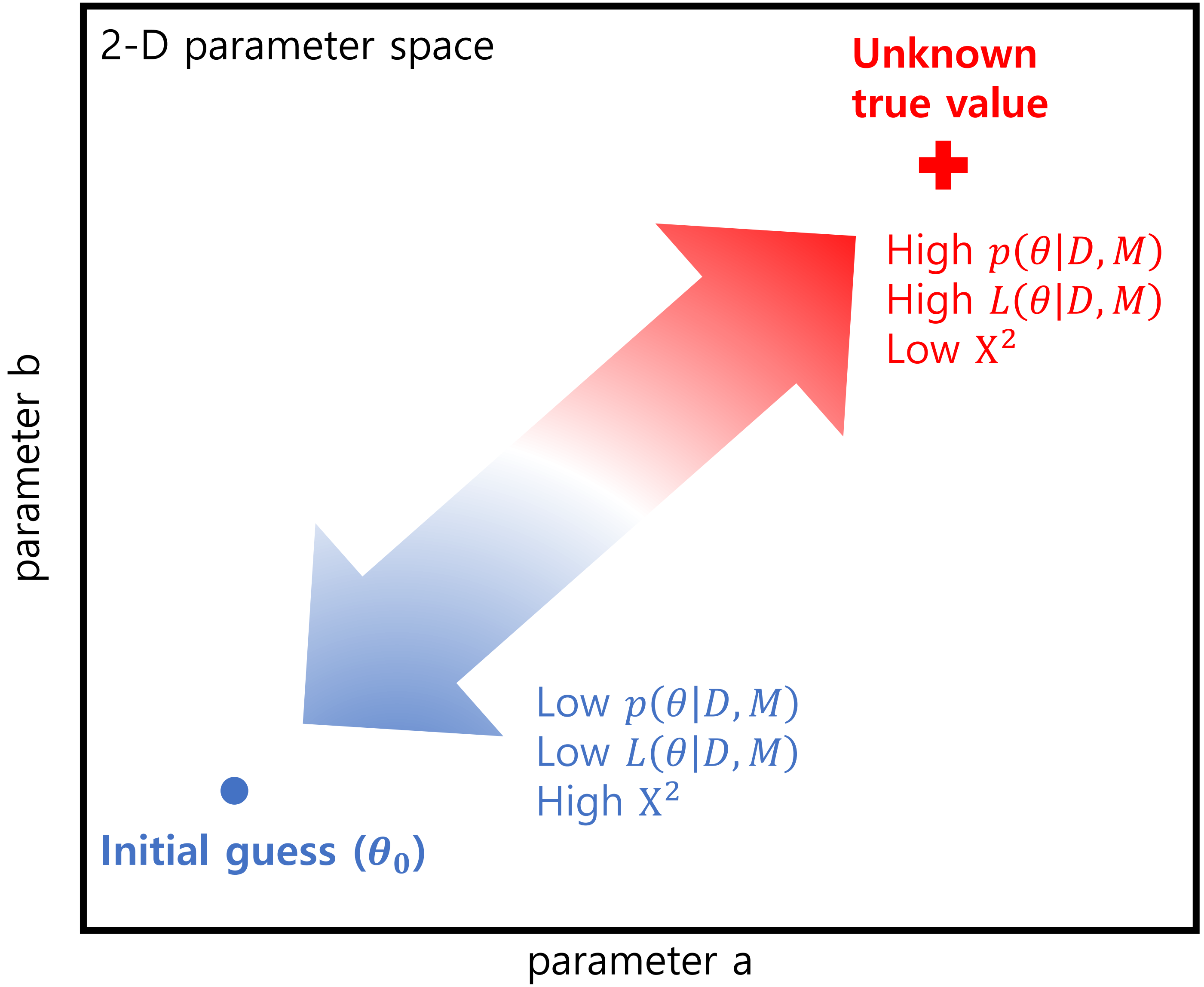
If $\mathrm{X}_{i+1}^2 \leq \mathrm{X}_i^2$, accept the jump

If $\mathrm{X}_{i+1}^2 > \mathrm{X}_i^2$, accept the jump with

$$p = \exp\left(-\frac{\Delta \mathrm{X}^2}{2}\right)$$

How to the MCMC Bayesian Estimation works?

- Example: 2-D problem
- $\theta = (a, b)$



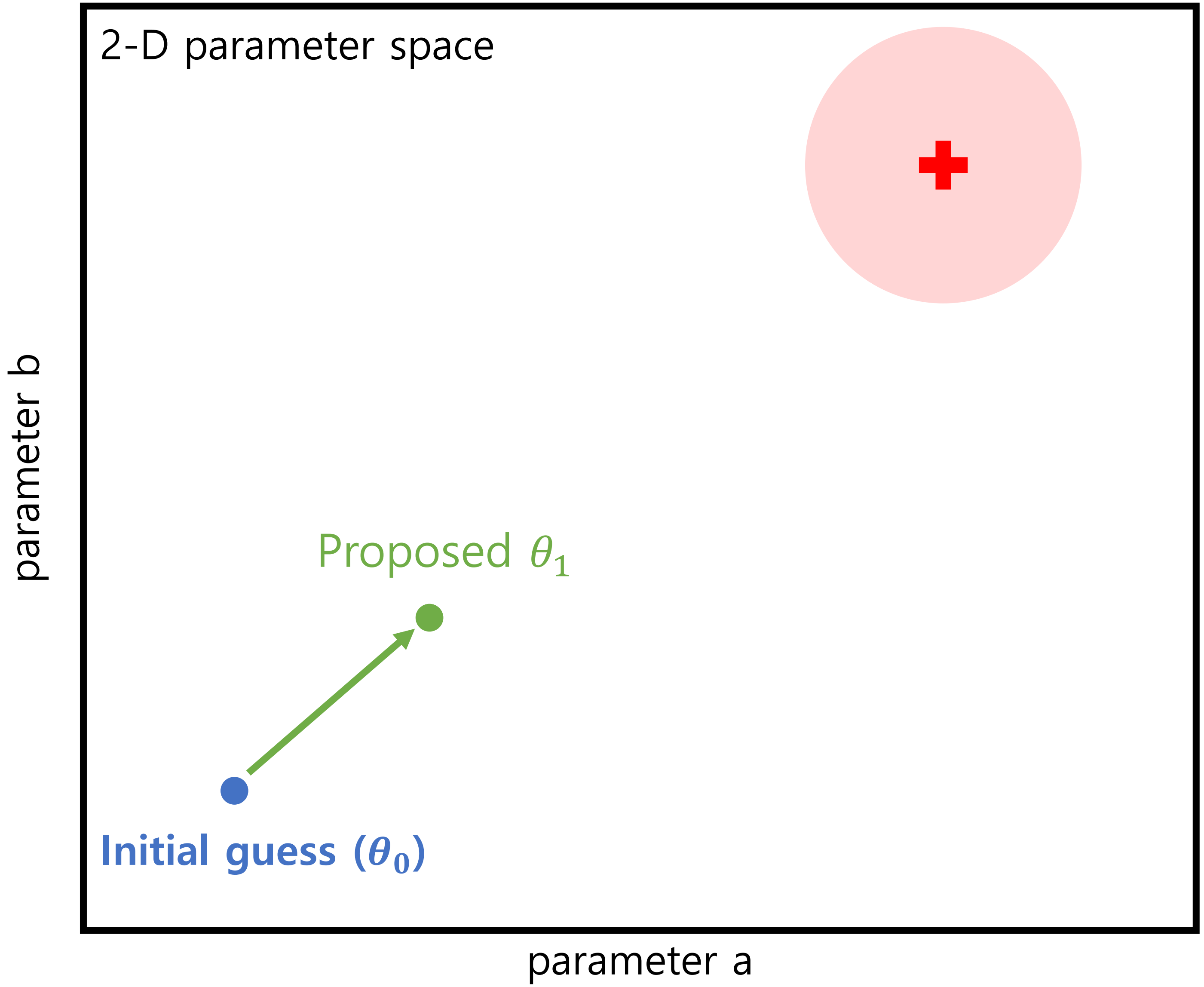
How to the MCMC Bayesian Estimation works?

- Example: 2-D problem
- $\theta = (a, b)$
- 새로운 step을 제안하기 위해 Normal distribution을 사용.

$$\theta_{i+1} = \theta_i + N(0, \Delta\theta)$$

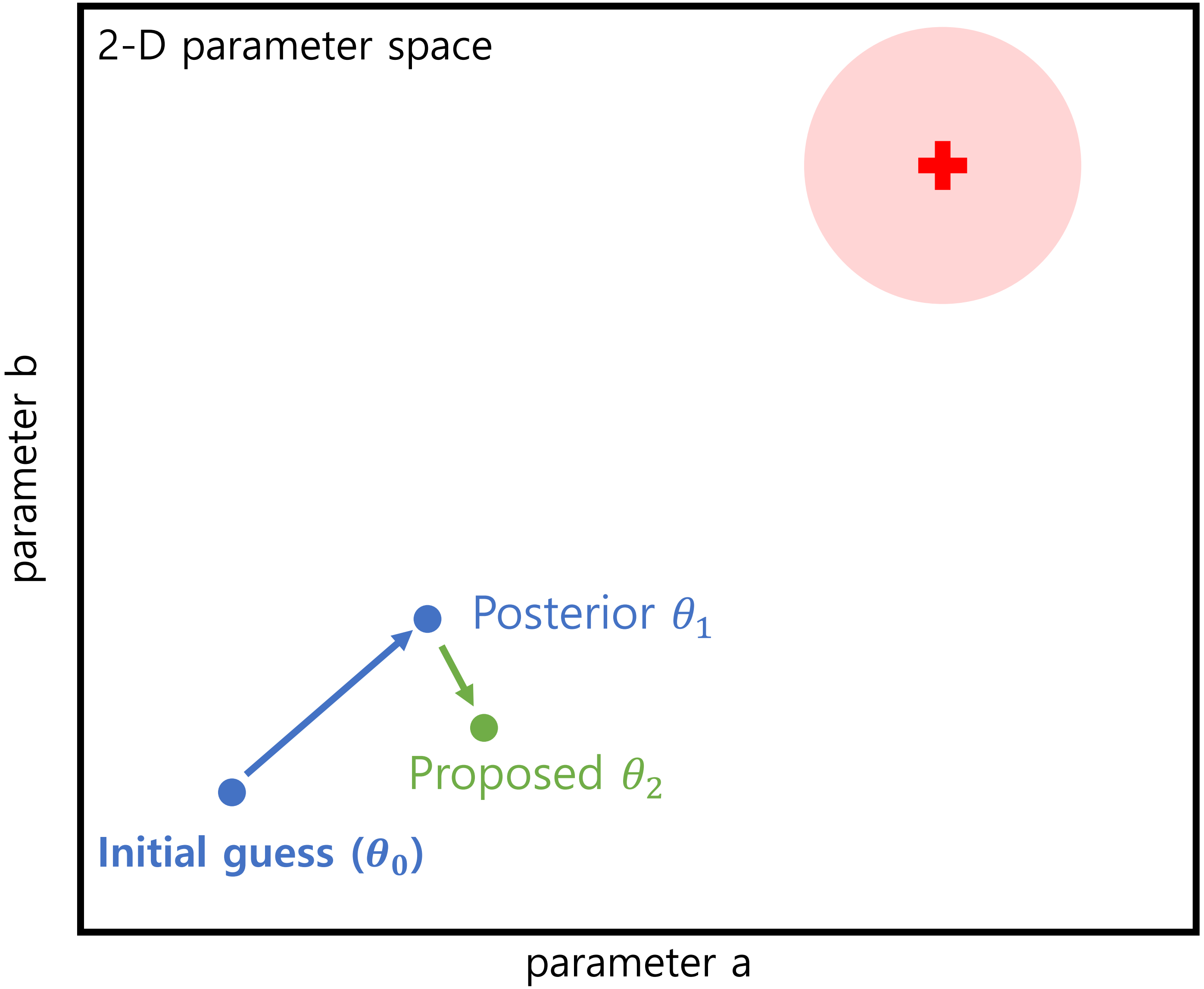
Jump scale

- Judge the acceptance



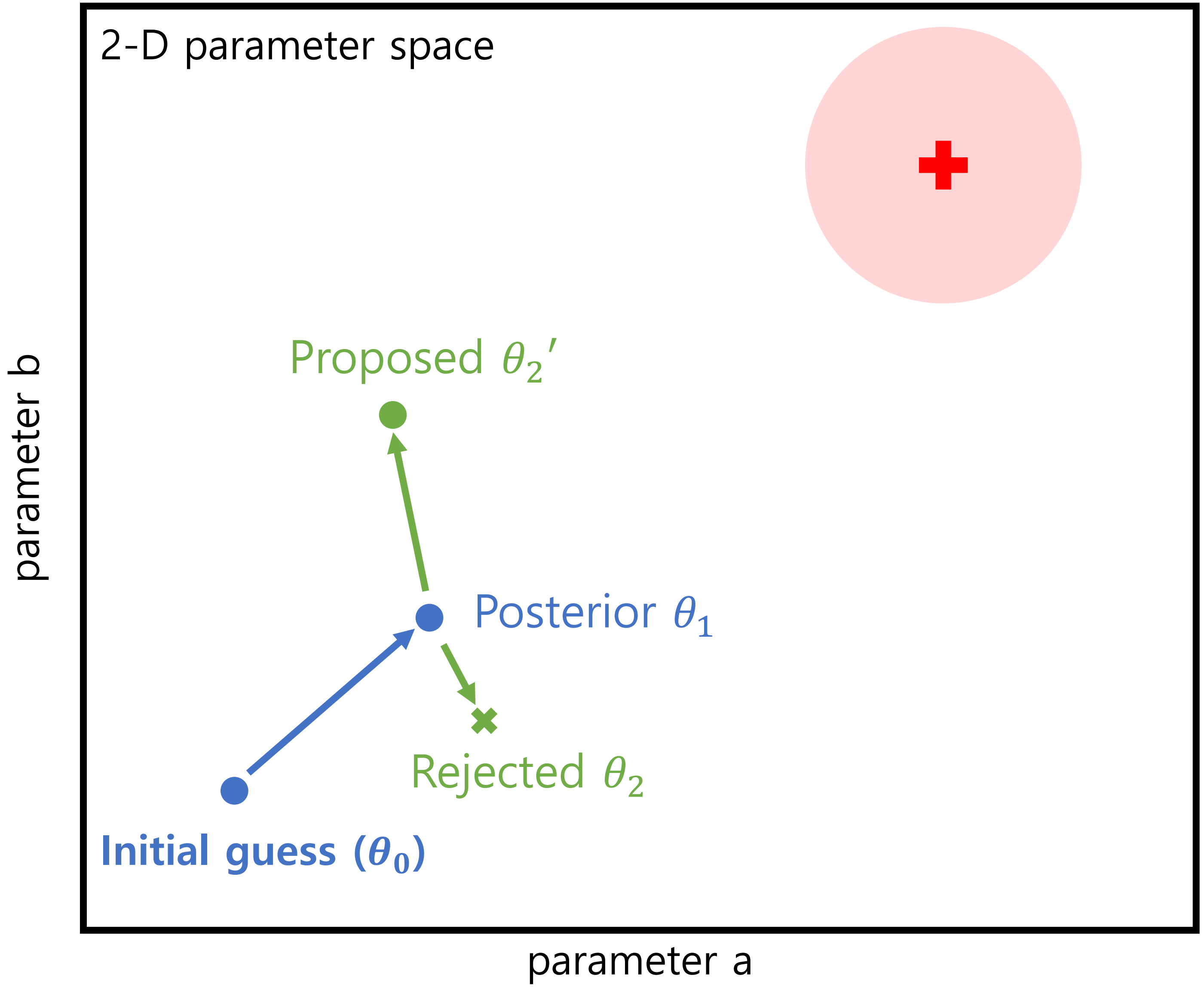
How to the MCMC Bayesian Estimation works?

- If θ_1 is accepted, propose the next step (θ_2)
- If the proposed step is rejected, propose another step ($\theta_2{}'$)
- Propose the next step until it is accepted.



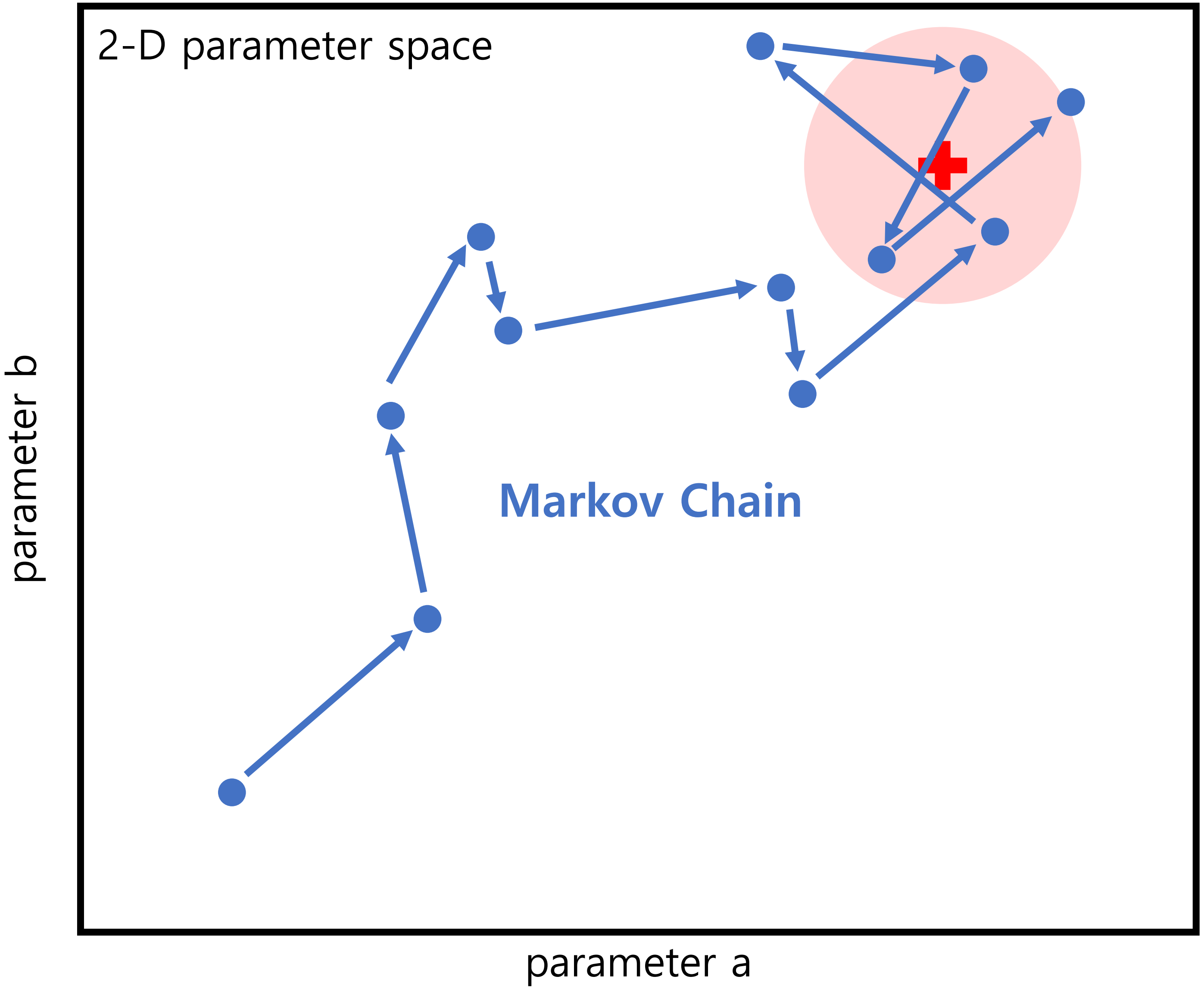
How to the MCMC Bayesian Estimation works?

- If θ_1 is accepted, propose the next step (θ_2)
- If the proposed step is rejected, propose another step (θ_2')
- Propose the next step until it is accepted.



How to the MCMC Bayesian Estimation works?

- Collect posteriors.
- The steps move around near the true values (orbit the solution and never stop at it).
- Keep jumping!

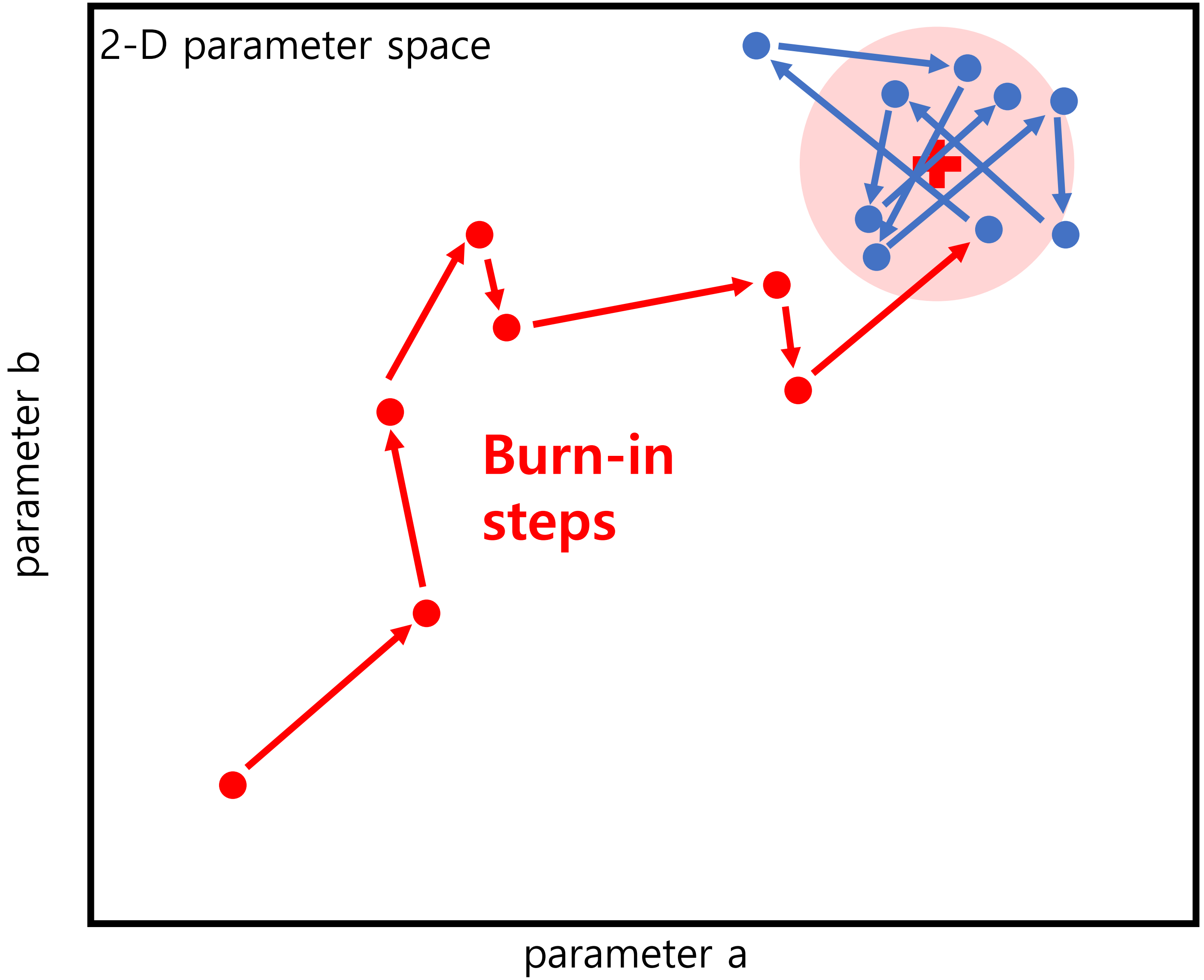


How to the MCMC Bayesian Estimation works?

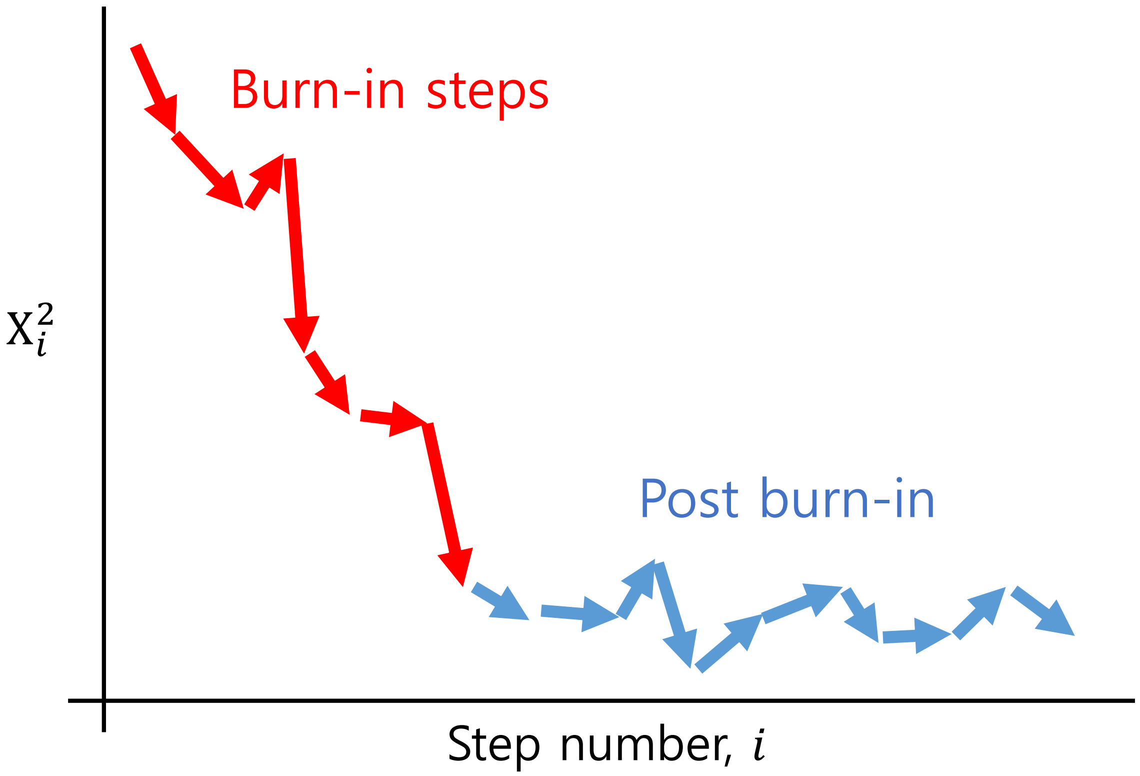
- Define Burn-in steps
- Remove burn-in steps

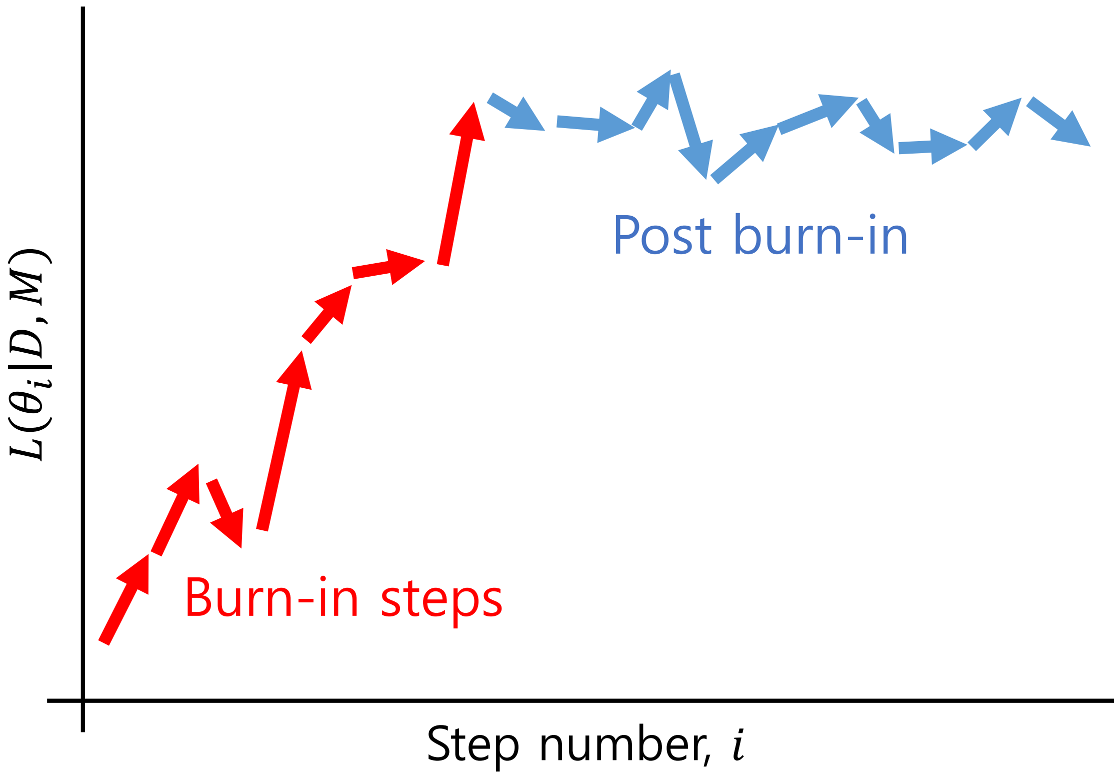
- Something checks to do
  1. Posteriors would be well mixed.
  2. The posteriors would converge on the same end point.

❖ Output: **the list of posteriors**



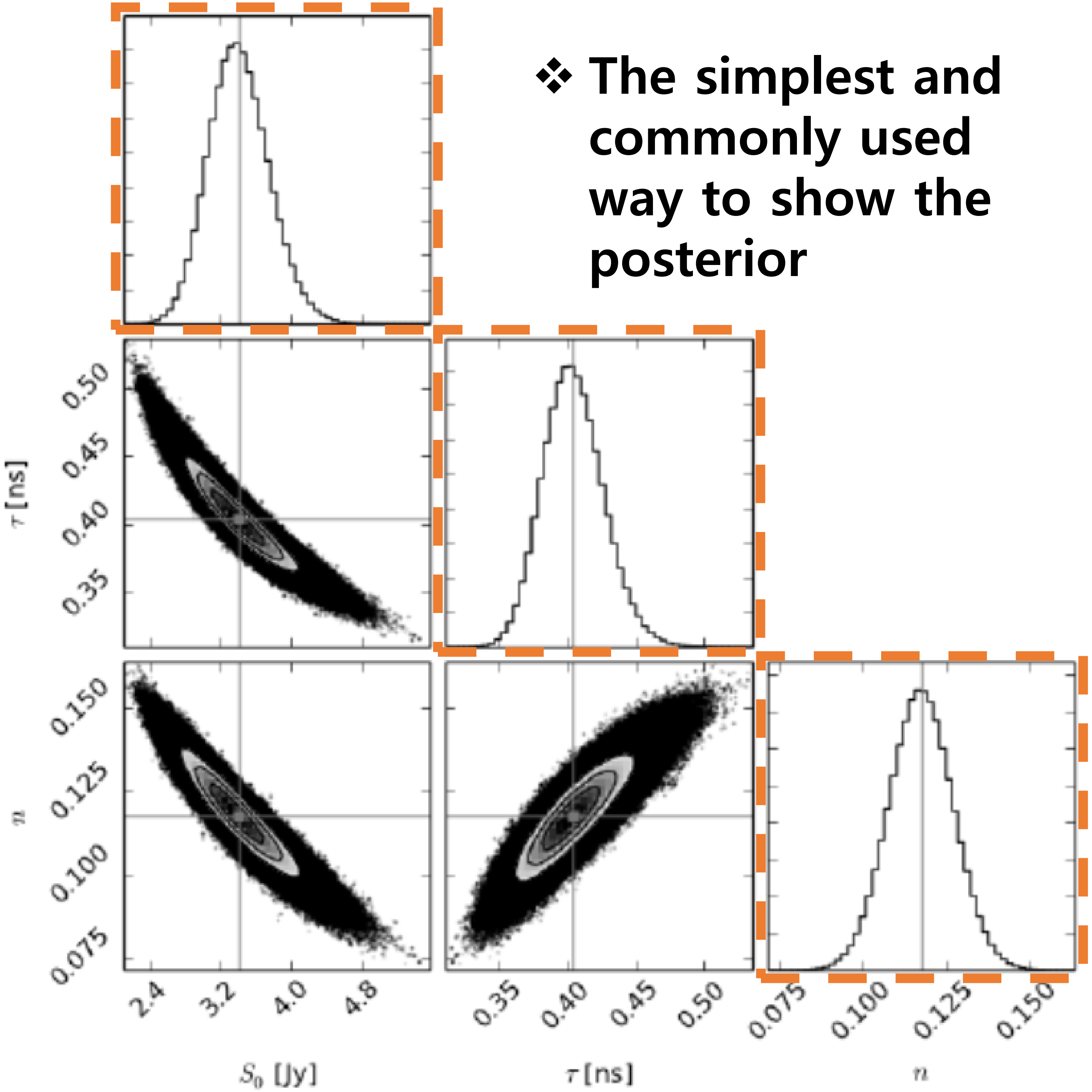
How to define the burn-in steps?





Triangle plot

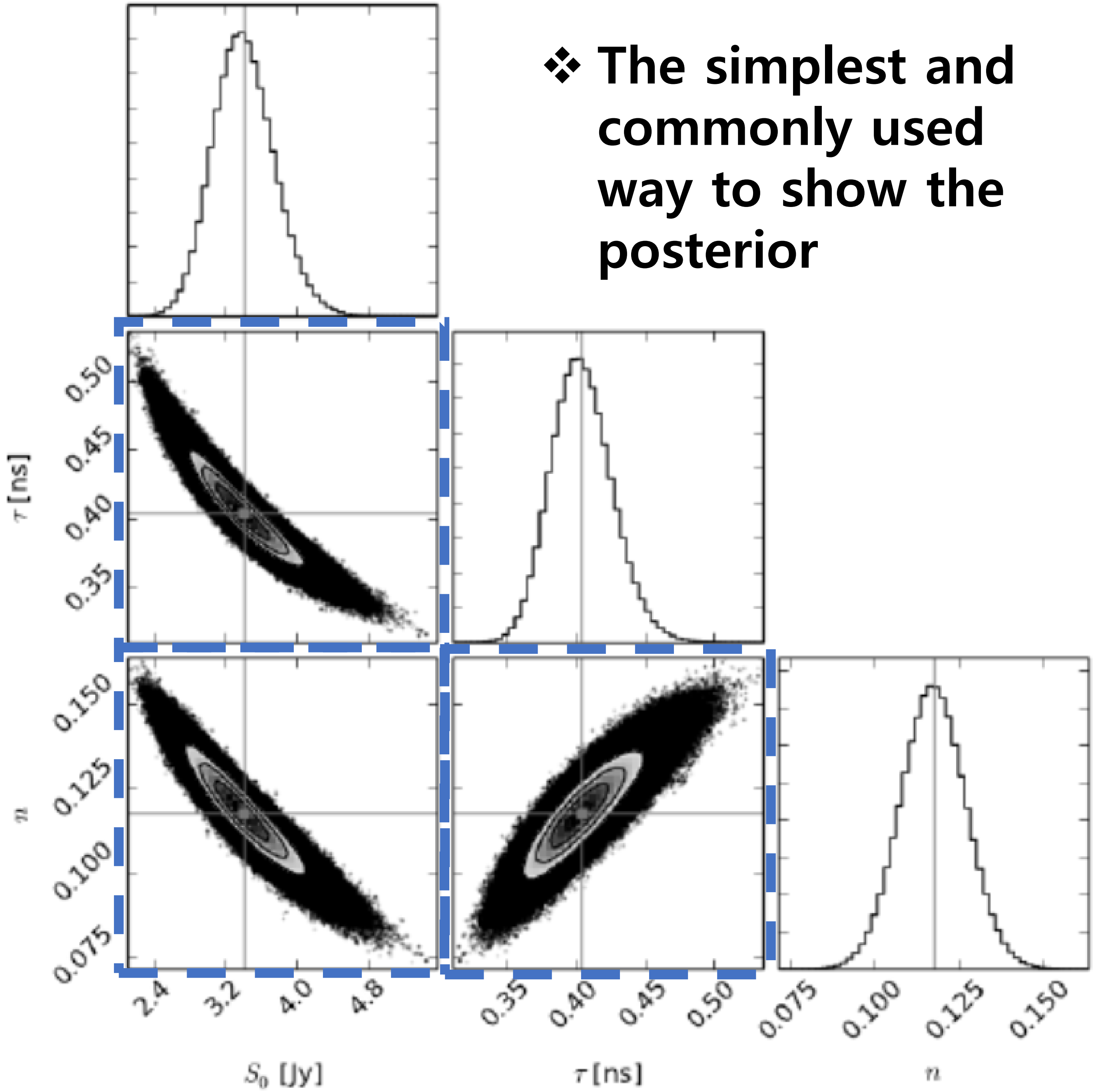
- **Histogram of each parameter**
  - Mean: the best-fit value of the parameter
  - Standard deviation: 1-σ error



❖ **The simplest and commonly used way to show the posterior**

Triangle plot

- **Histogram of each parameter**
  - Mean: the best-fit value of the parameter
  - Standard deviation: 1-σ error
- **Posterior distribution**
  - The joint probability distribution between two parameters
  - We can check the correlation between the parameters

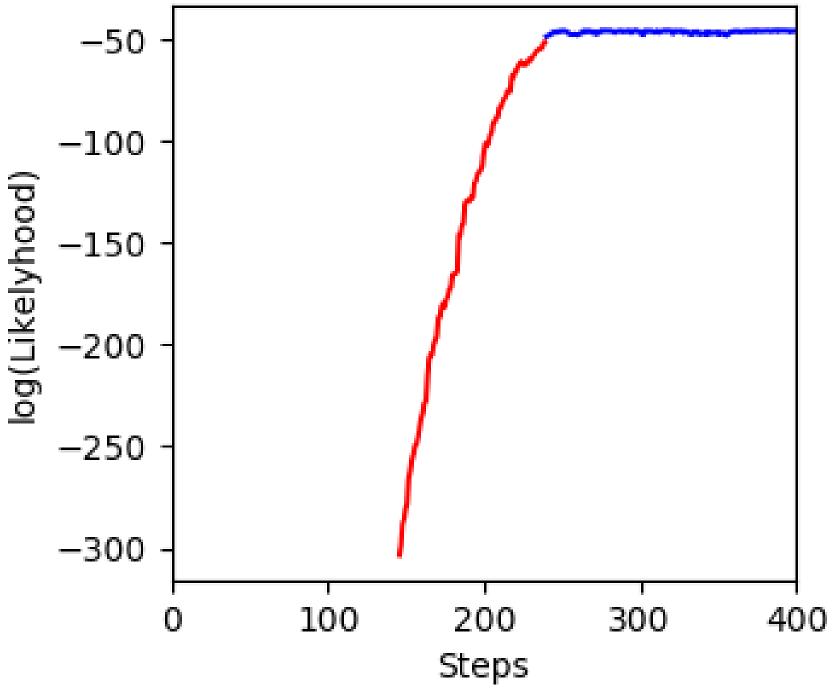


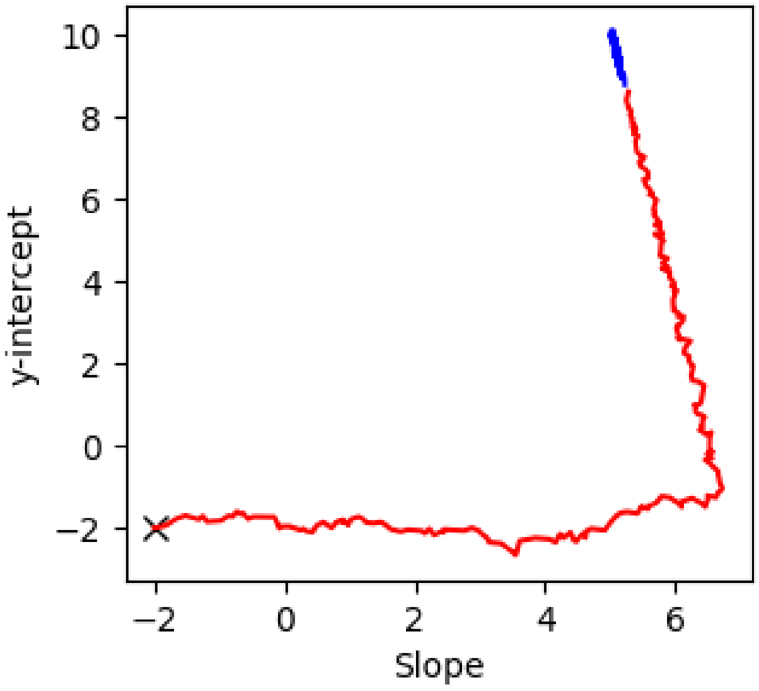
❖ **The simplest and commonly used way to show the posterior**

Test problem 1.

- Linear fitting using the MCMC Bayesian estimation code.

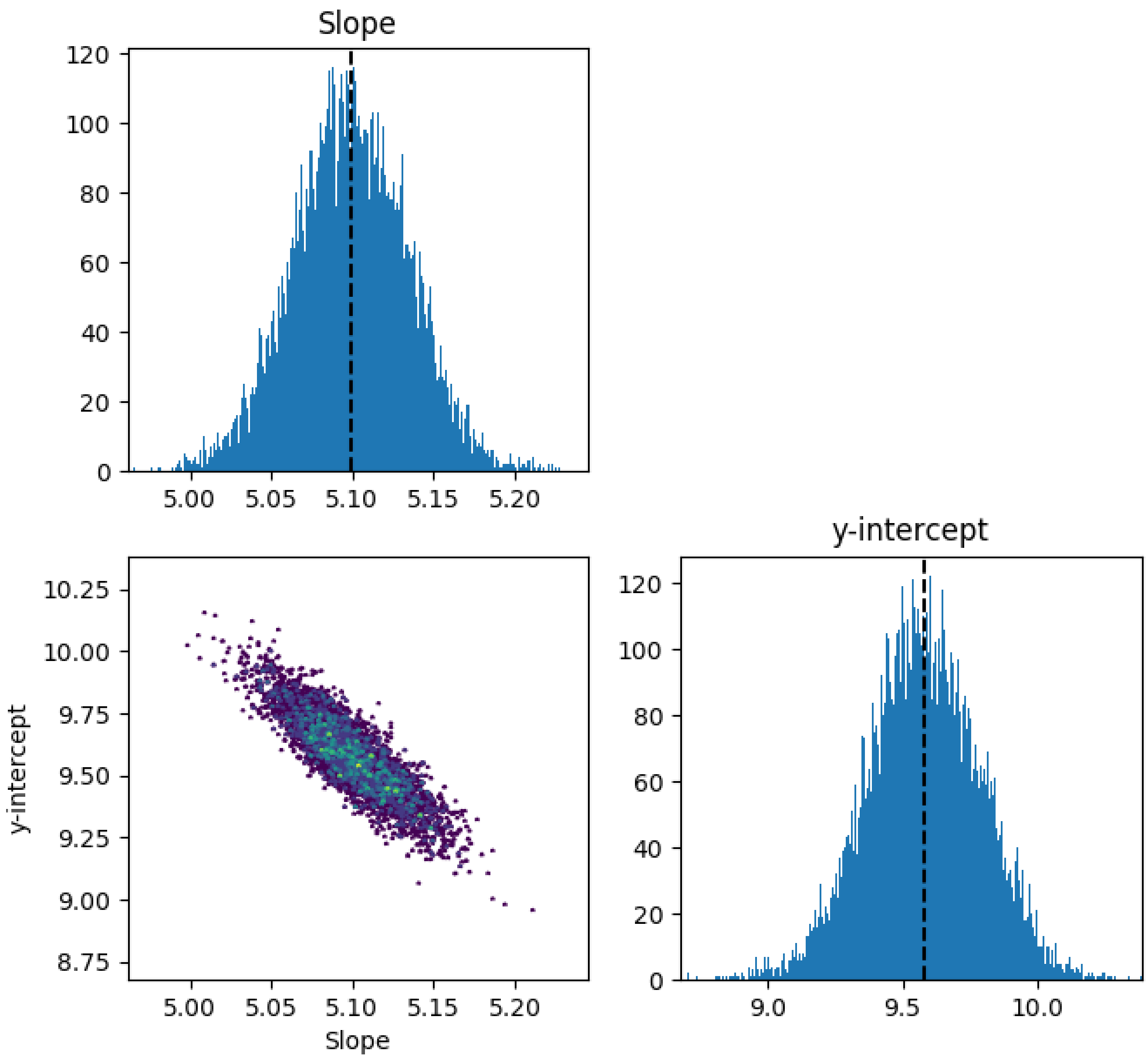
  [HyeongSikYun/MCMC_linear_fitting: the linear fitting code using MCMC (github.com)]





Test problem 1.

- Triangle plot



Test problem 2.

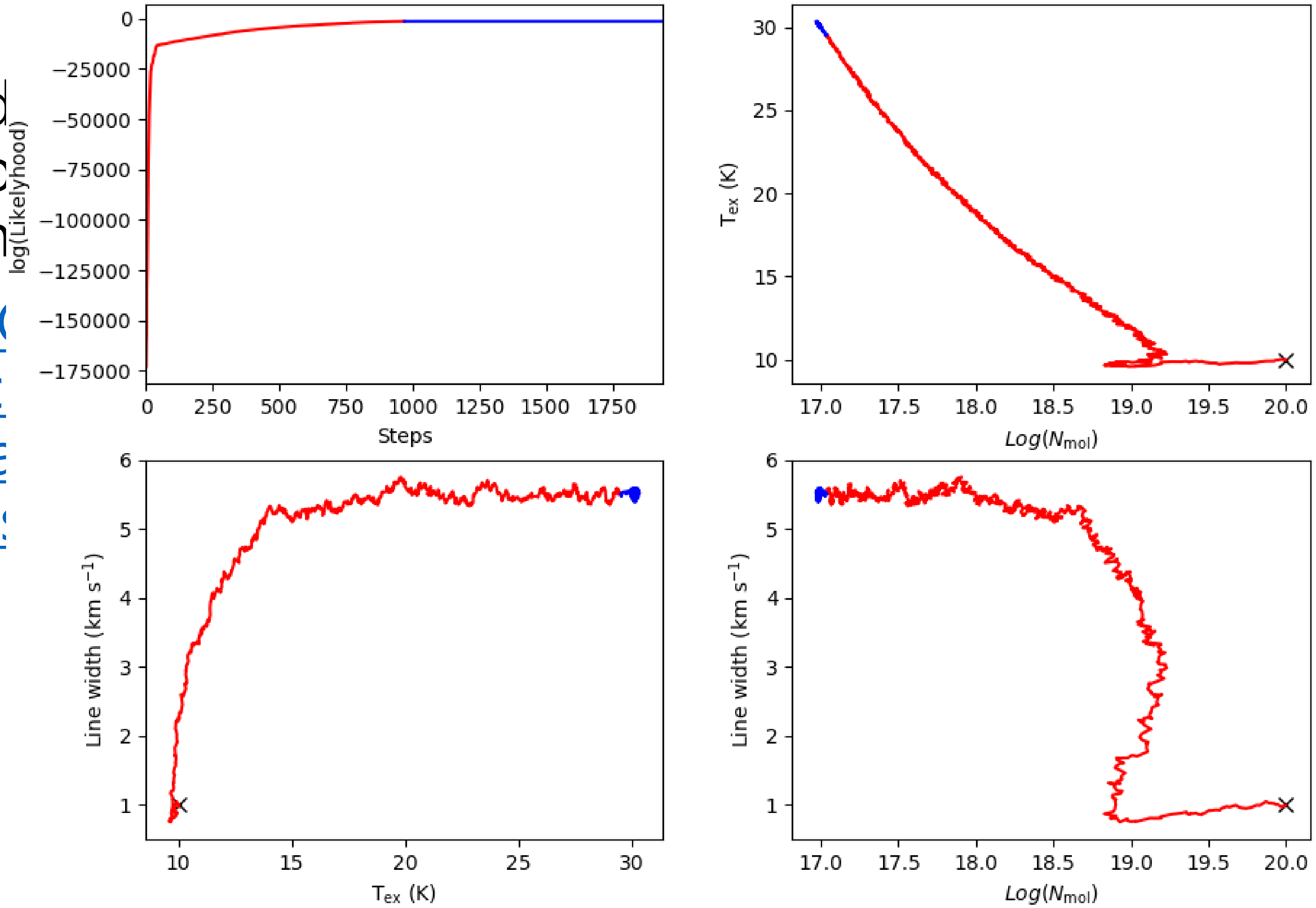
- Derive the column density and excitation temperature of molecule from the observed ALMA spectrum using the rotational diagram model.

HyeongSikYun/MCMC_line_fitting: This is an application of the MCMC linear fitting code to derive the gas column density and excitation temperature from the observed molecular line spectra. (github.com)

Test problem 2.

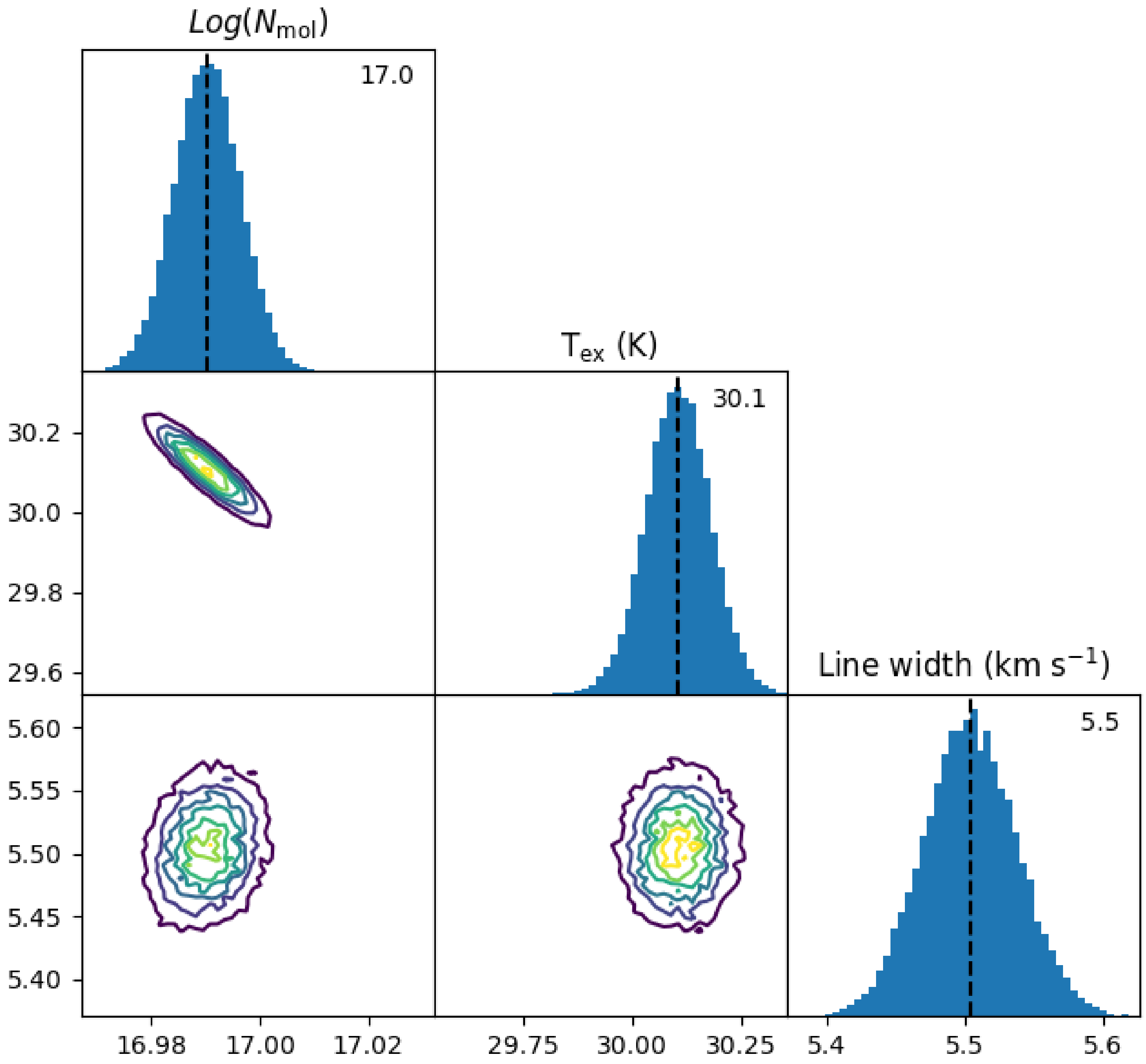
- Derive the column d
molecule from the c
rotational diagram n

HyeongSikYun/MCM(
MCMC linear fitting c
excitation tempera



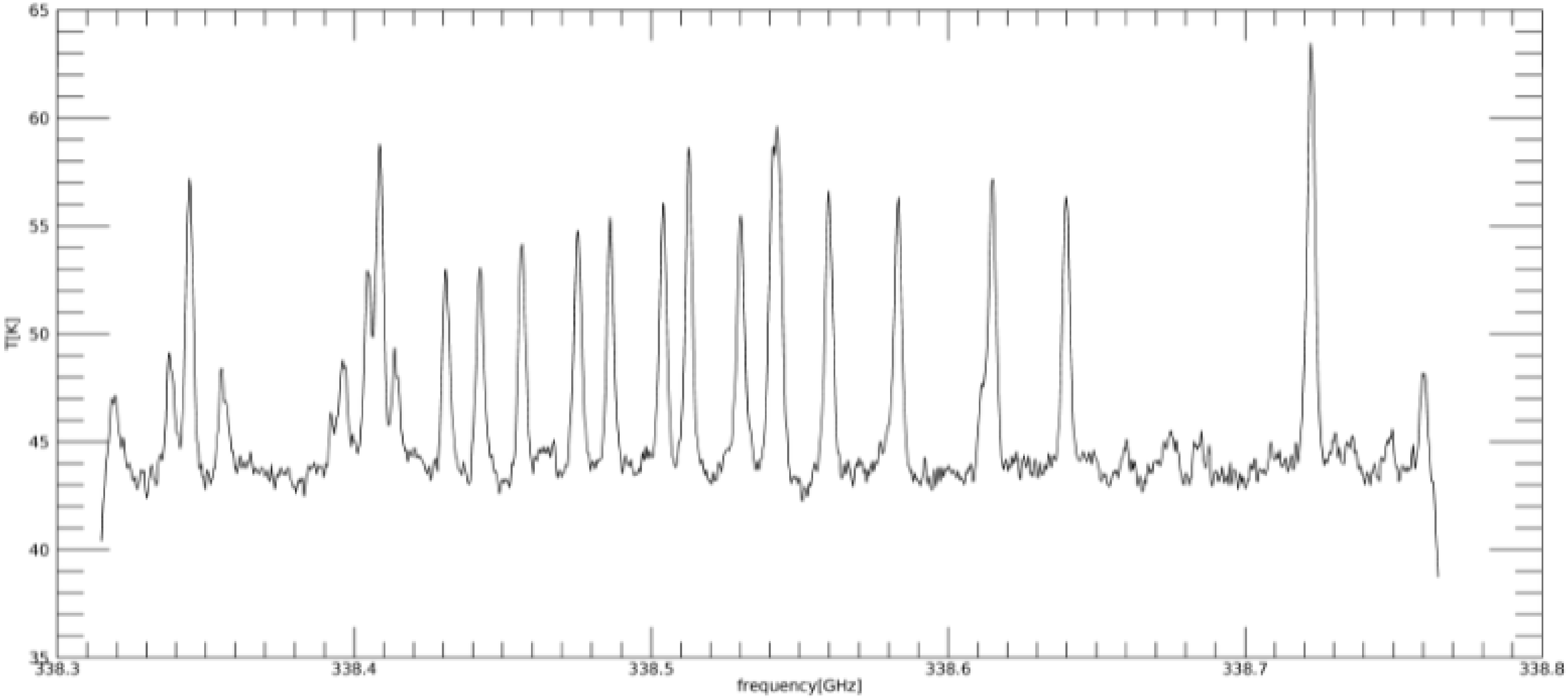
Test problem 2.

- Triangle plot



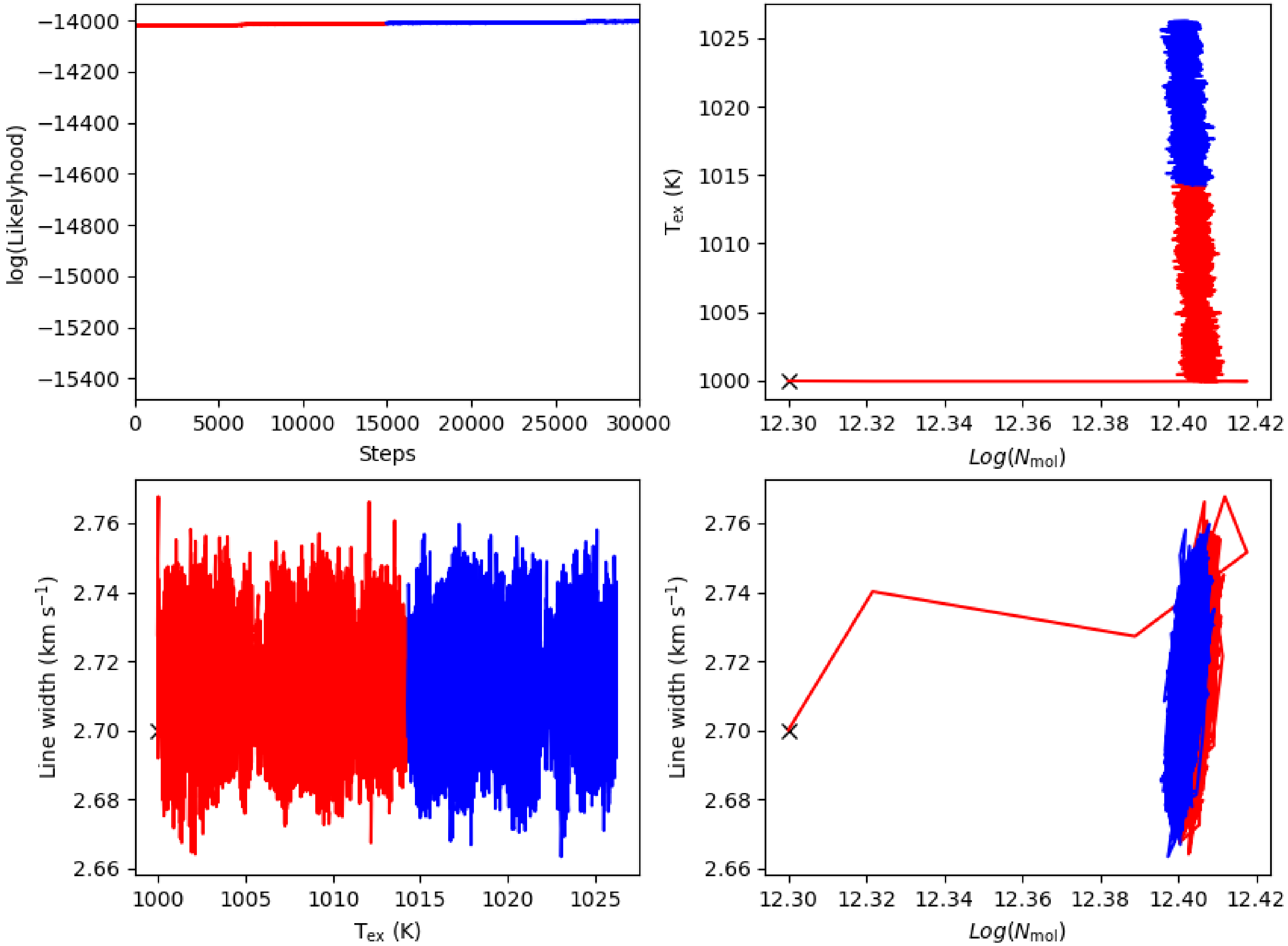
Application of test problem 2.

- Try to fit the observed ALMA spectrum
- Data: Methanol line spectrum toward V883 Ori



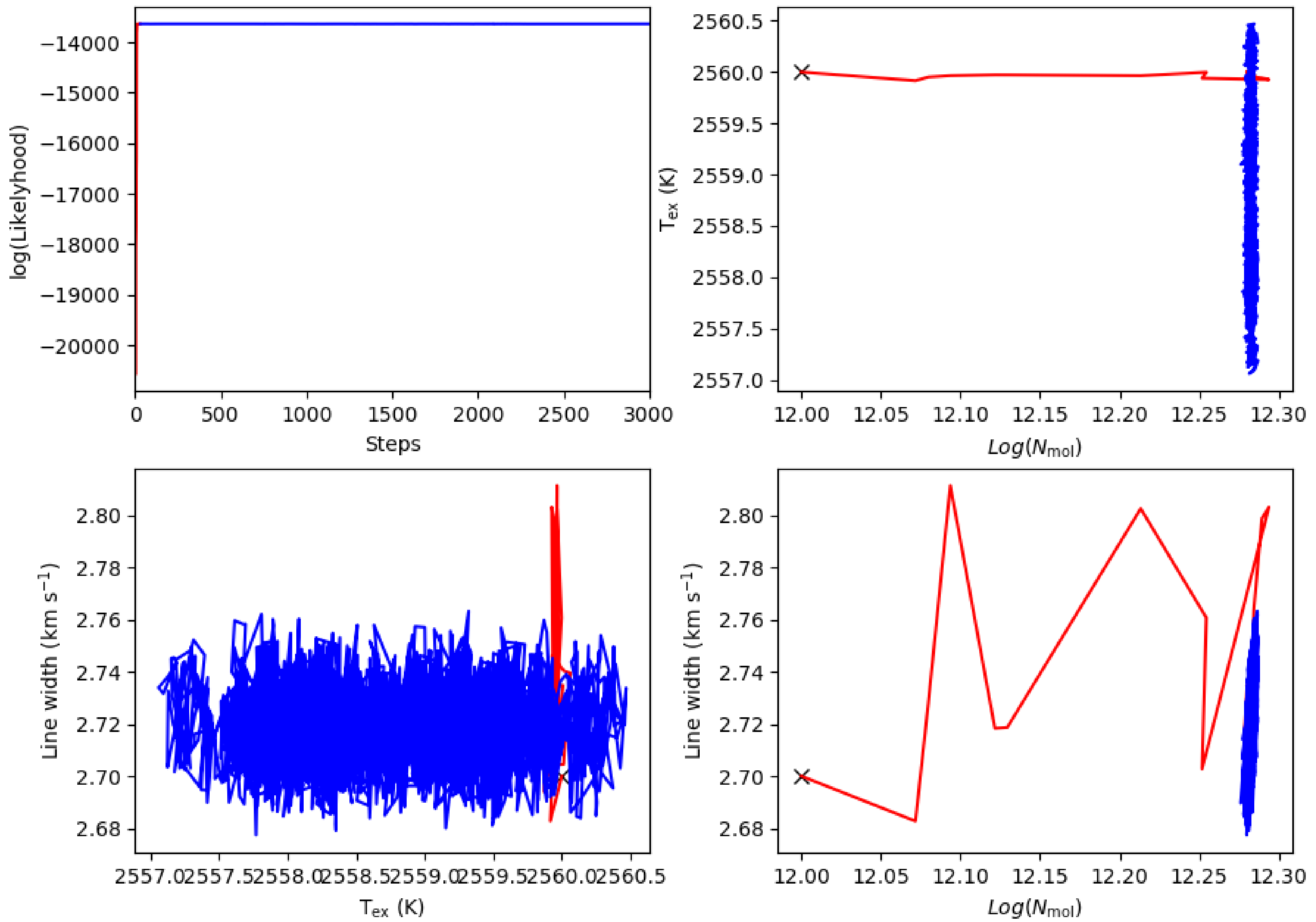
Application of test problem 2.

- Did not converge



Application of test problem 2.

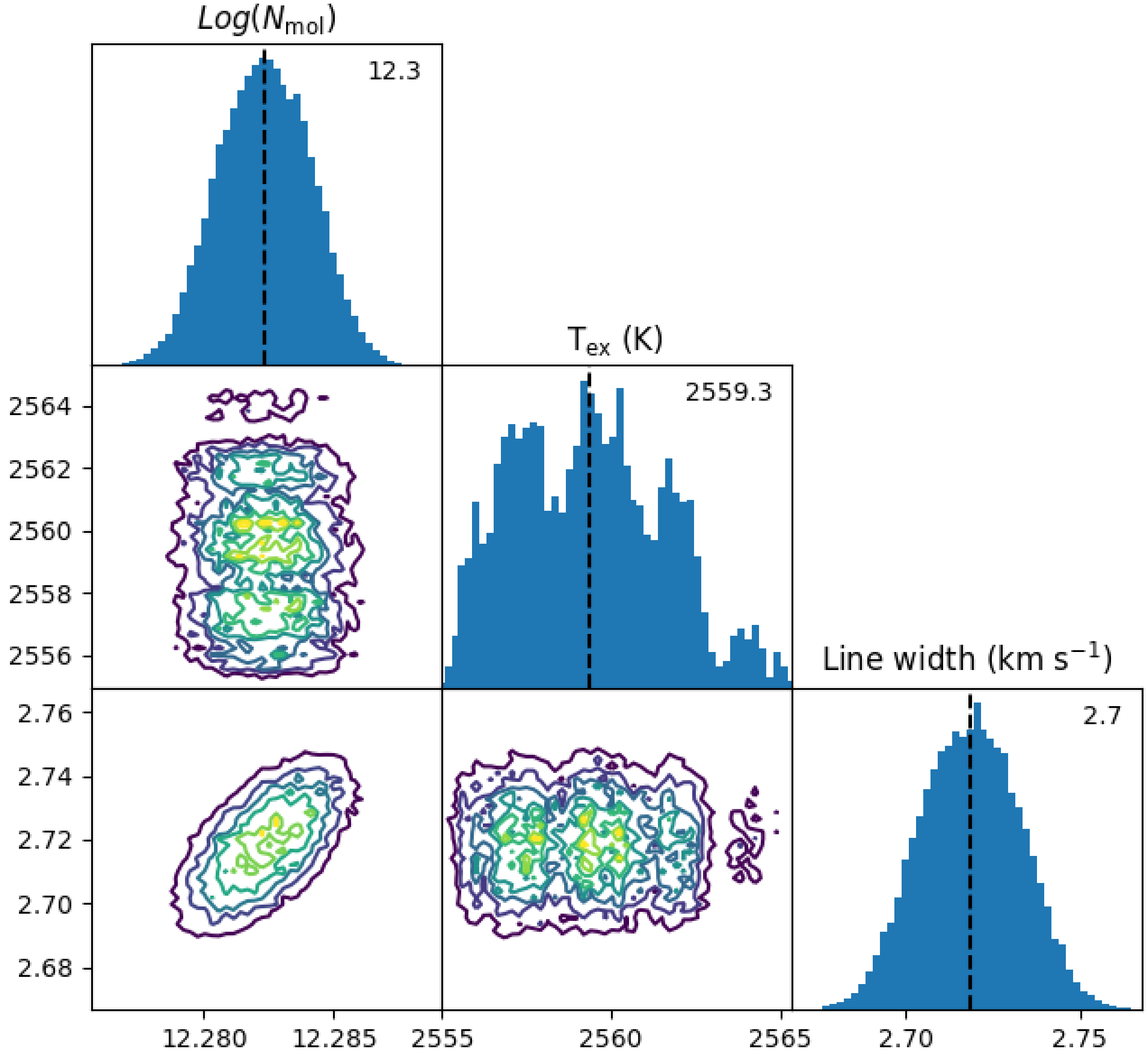
- Finally I found a end point (guess)



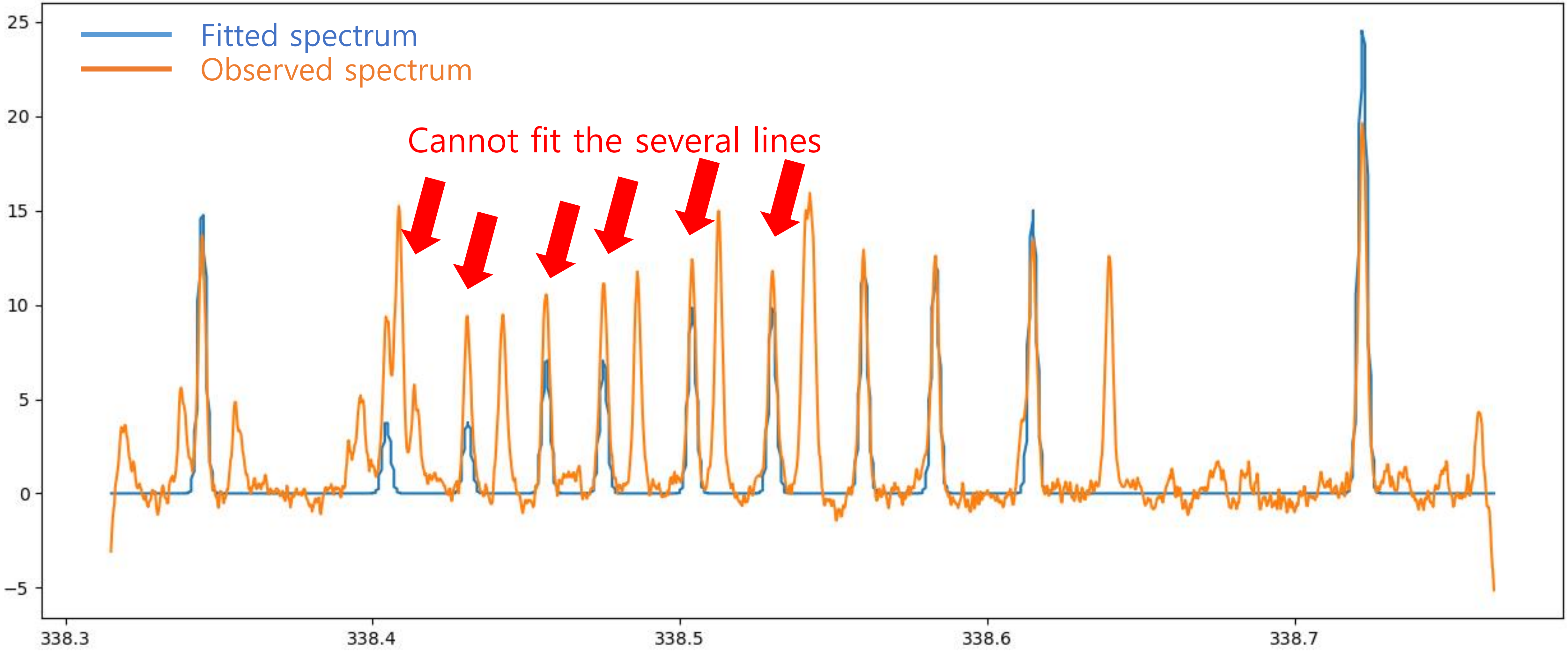
Application of test problem 2.

- Finally I found a end point (guess)
- The excitation temperature is too high.
- The expected thermal width exceeds the fitted line width.

❖**Not the solution.**



Application of test problem 2.

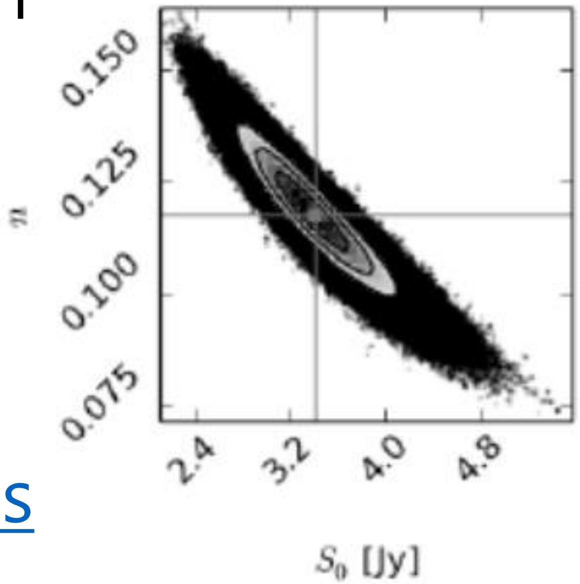


Advanced application

- 구하려는 parameter 사이에 correlation이 없는 것이 가장 좋다.
- 하지만 parameter 사이에 correlation이 있는 경우가 있고, 이 경우 MCMC가 진행해 나가기 힘들다.
- 이를 고려한 더 Fancy한 algorithm이 있다.

❖Ex) Gaussian process likelihood

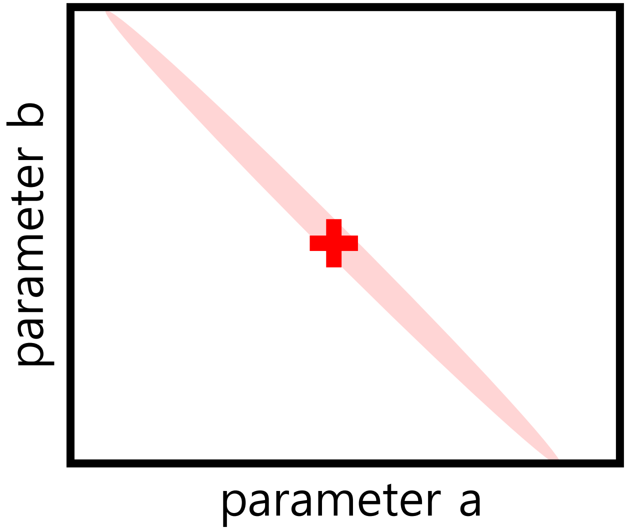
[An Astronomer's Introduction to Gaussian Processes - Speaker Deck]



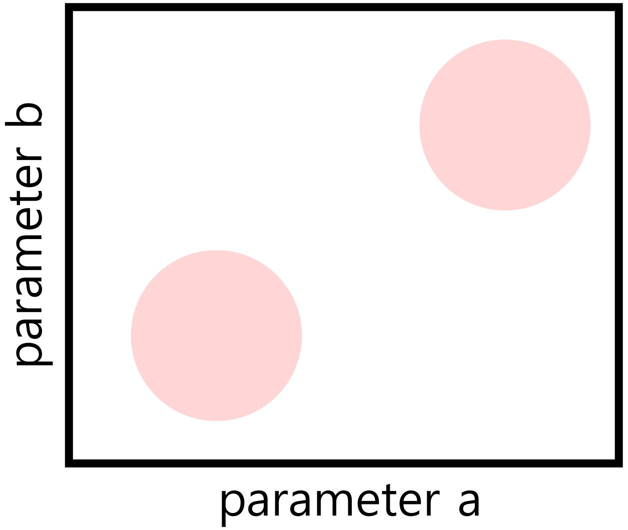
Advanced application

- 관측 noise가 Normal white noise로 설명되지 않을 수 있다.
  경우에 따라 원하는 likelihood function을 사용해야 한다.

- Metropolis rule이 빠르고 정확하게 작동하지 못하는 경우도 있다.



or



- 다양한 방법들을 사용
  1. Metropolis-Hasting rule
  2. Simulated Annealing
  3. Parallel tempering
  4. Affine-invariant sampling

References

- MCMC class in Sagan Summer Workshop 2016

  A Beginner's Guide to Monte Carlo Markov Chain MCMC Analysis 2016 - YouTube
- Youtube channel (공돌이의 수학정리노트)

  Markov Chain Monte Carlo(1): Sampling편 - YouTube

  Markov Chain Monte Carlo(2): Bayesian Estimation – YouTube

  베이즈 정리의 의미 – YouTube

  최대우도법(Maximum Likelihood Estimation) 소개 – YouTube
- MCMC 내용 & 예제

  Markov Chain Monte Carlo - 공돌이의 수학정리노트 (angeloyeo.github.io)