
Probabilistic Machine Learning:

10. Sampling Methods

ESC 2024 Spring Session 4주차



Contents

1. Basic Sampling Algorithms

2. MCMC

3. Gibbs Sampling

4. Slice Sampling

5. Hybrid Monte Carlo Algorithm

6. Estimating the partition Function

1

Basic Sampling Algorithms

Standard distribution
Importance sampling

Basic Sampling Algorithms

실제로 활용하는 확률적 모델중 많은 것들은 정확한 추론을 직접 시행하기가 매우 까다롭다. 이런 경우 근사치를 사용한다

10장에서 결정적근사를 기반으로한 추론 알고리즘 ex. 변분적베이지스/EP

11장에서 수치적표집법을 기반으로한 근사추론법 Monte Carlo 테크닉을 배울 것이다.

<목표>

사후분포에 대한 기댓값이 중요하지만 복잡한 함수일수록 계산의 어려움을 겪는다.

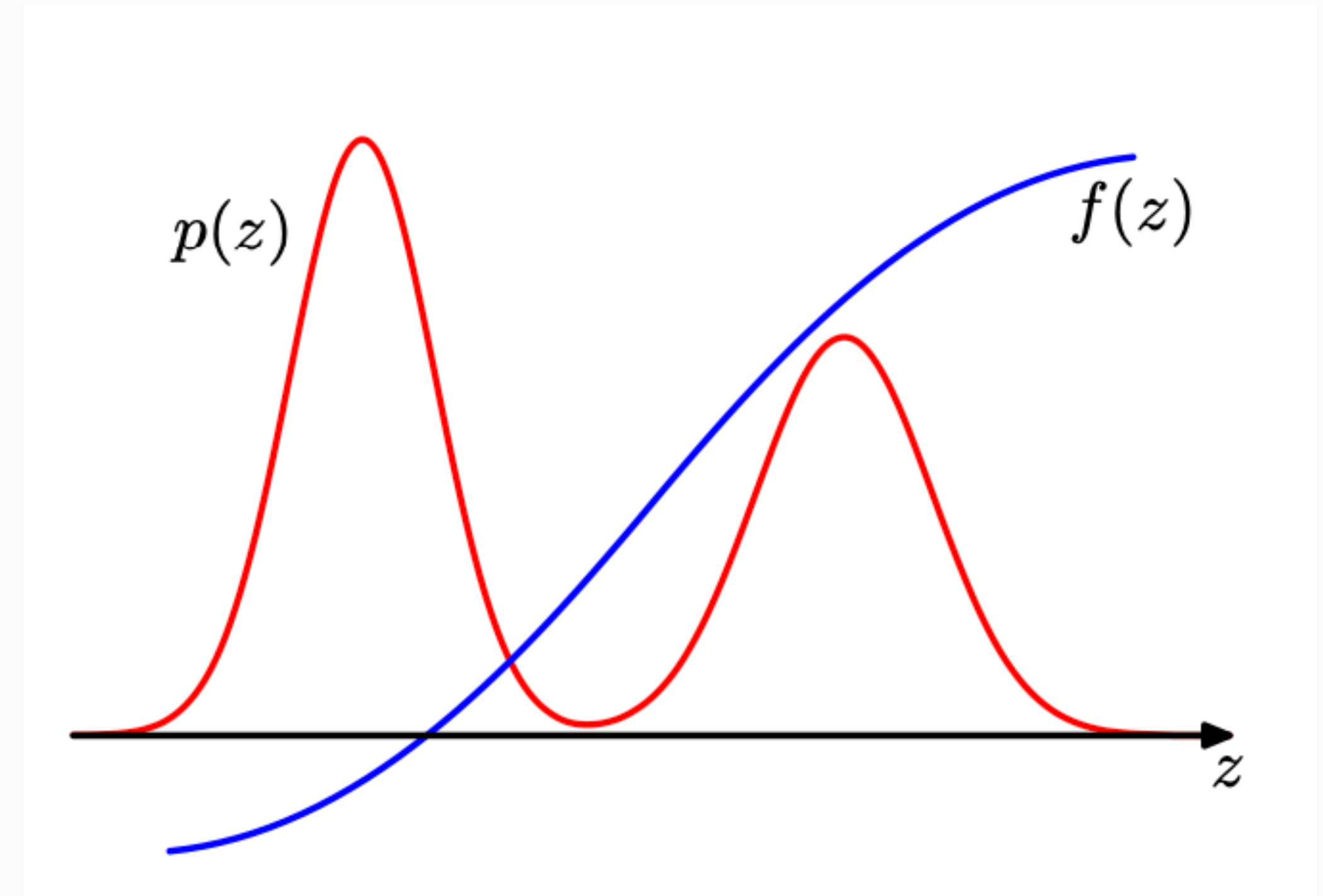
$$\mathbb{E}[f] = \int f(\mathbf{z})p(\mathbf{z}) d\mathbf{z}$$

Basic Sampling Algorithms

<표집법의 기본아이디어>

$$\hat{f} = \frac{1}{L} \sum_{l=1}^L f(\mathbf{z}^{(l)}).$$

- 표집법을 통한 근사값은 큰수의 법칙에 근거한다.
- 추정량의 정확도는 z 의 차원수에 종속적이지 않고 상대적으로 적은 수의 표본들로부터도 높은 정확도를 달성 할 수 있다.
- 하지만 표본끼리 독립적이지 않을때 유효표본의 크기가 실패표본의 크기보다 훨씬 작을 수 있고, 높은 정확도를 얻기 위해 상대적으로 많은 수의 표본이 필요하다.



Basic Sampling Algorithms (Standard distributions)

목표: Y에서 랜덤한 수 생성

1) $y = f(z)$

- z is uniformly distributed $(0, 1)$

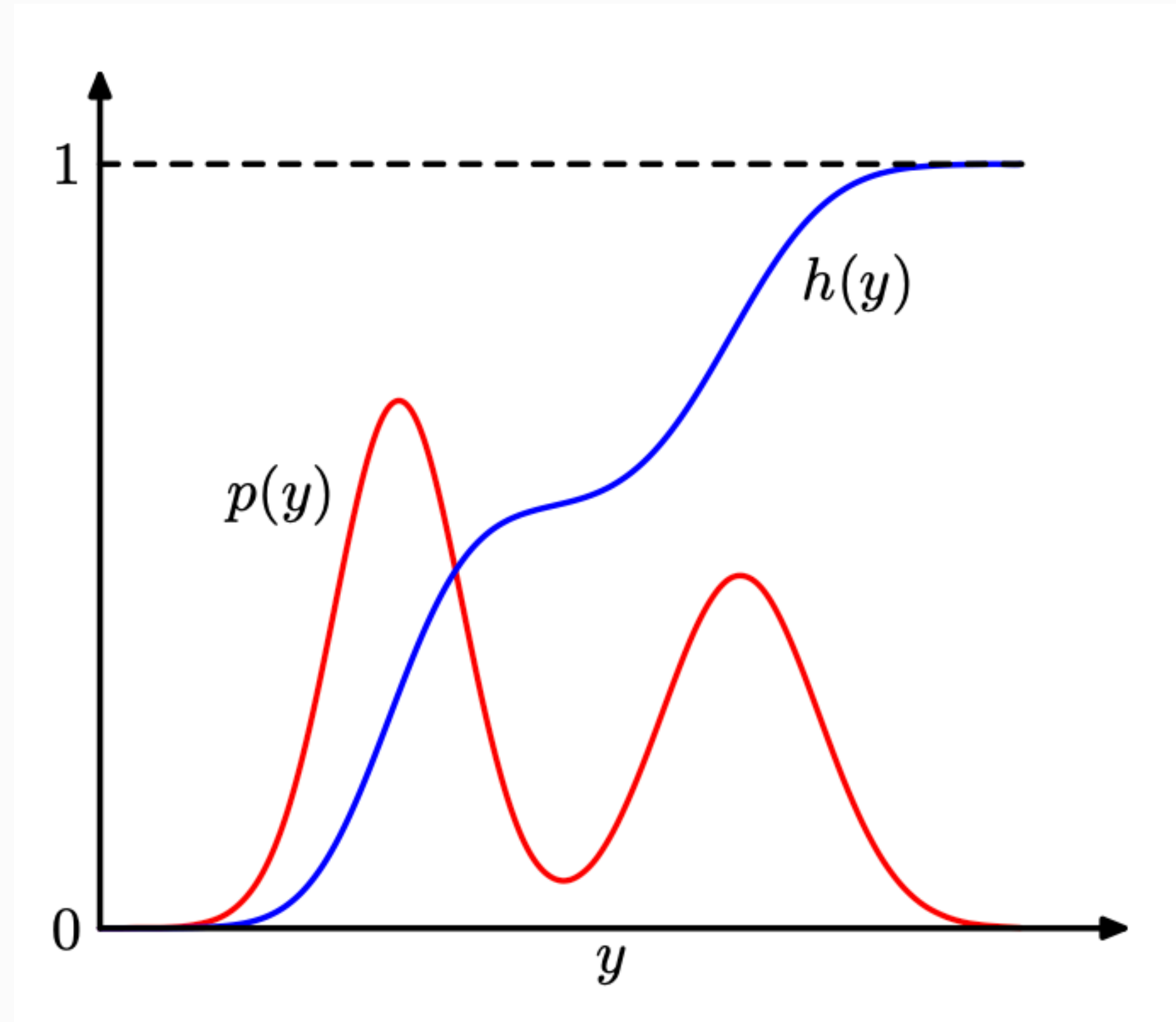
$$p(y) = p(z) \left| \frac{dz}{dy} \right|$$

2) 적분(cdf 생성)

$$z = h(y) \equiv \int_{-\infty}^y p(\hat{y}) d\hat{y}$$

3) 역함수 계산

$$y = h^{-1}(z)$$



Basic Sampling Algorithms (Standard distributions)

지수분포 예시

1)

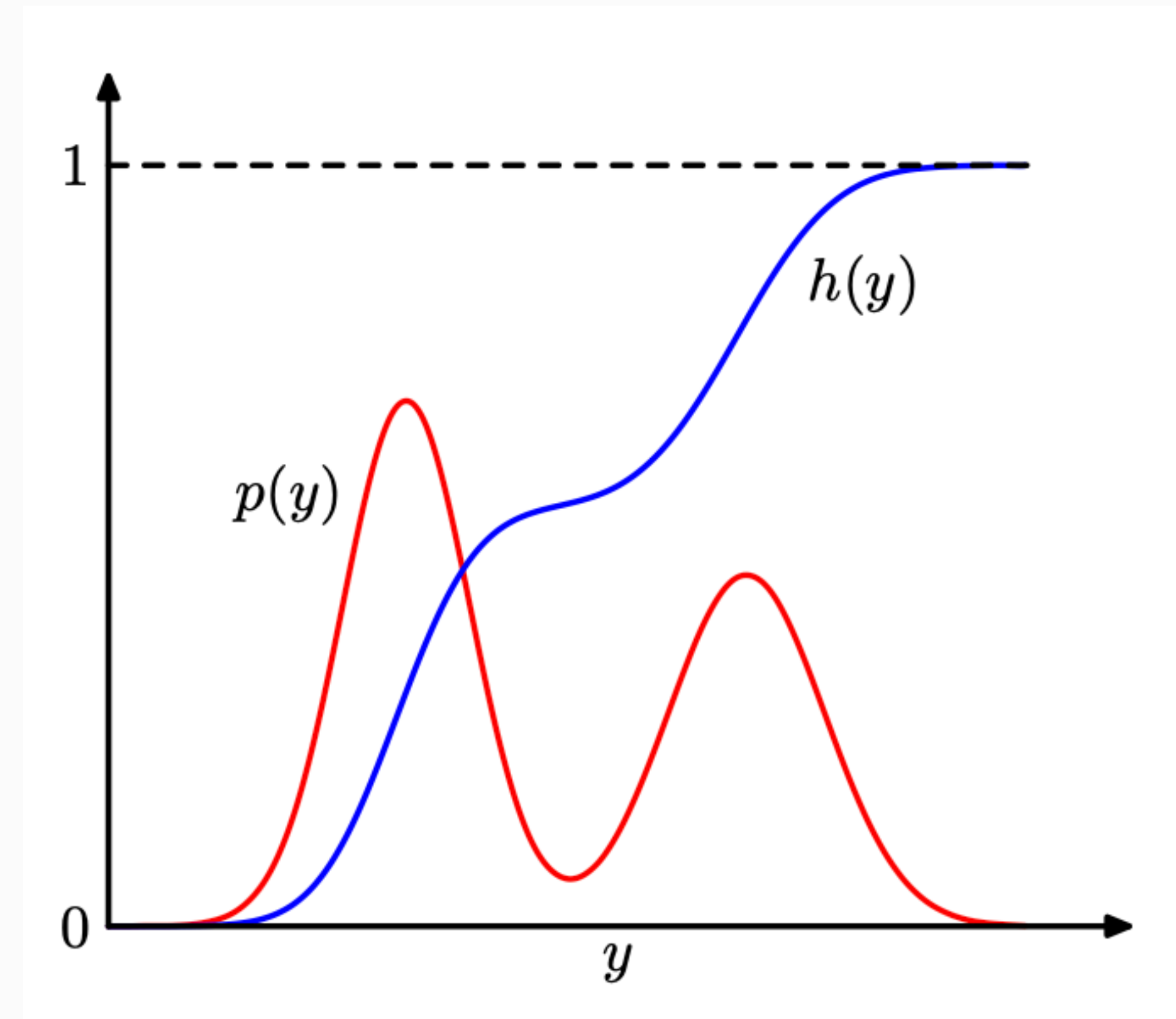
$$p(y) = \lambda \exp(-\lambda y)$$

2) cdf계산

$$h(y) = 1 - \exp(-\lambda y)$$

3) cdf 역함수 계산

$$y = -\lambda^{-1} \ln(1 - z).$$

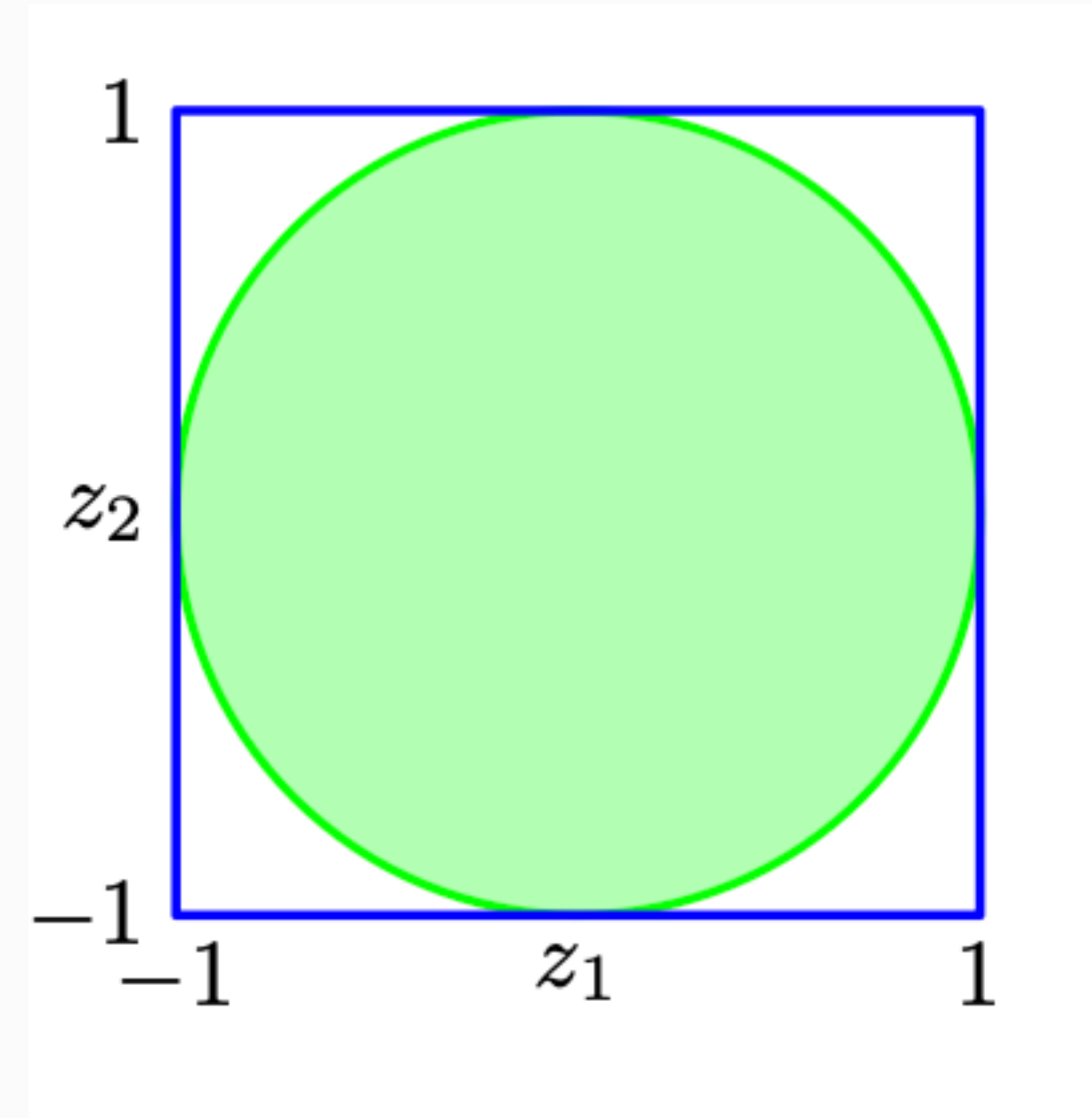


Basic Sampling Algorithms (Standard distributions)

가우시안분포 Box-Muller 방법

1) 단위 원안의 균등분포로부터 표본의 생성

- $(-1,1)$ 균일 분포에서 랜덤한 순서쌍 z_1, z_2 생성
- $z_1^2 + z_2^2 > 1$ 쌍 제거
- 결과적으로 원내부 $p(z_1, z_2) = 1/\pi$ 균등하게 분포



Basic Sampling Algorithms (Standard distributions)

가우시안분포 예시

2) y_1, y_2 확률변수 정의

$$\begin{aligned} y_1 &= z_1 \left(\frac{-2 \ln z_1}{r^2} \right)^{1/2} \\ y_2 &= z_2 \left(\frac{-2 \ln z_2}{r^2} \right)^{1/2} \end{aligned}$$

3) 결합분포 계산

$$\begin{aligned} p(y_1, y_2) &= p(z_1, z_2) \left| \frac{\partial(z_1, z_2)}{\partial(y_1, y_2)} \right| \\ &= \left[\frac{1}{\sqrt{2\pi}} \exp(-y_1^2/2) \right] \left[\frac{1}{\sqrt{2\pi}} \exp(-y_2^2/2) \right] \end{aligned}$$

4) 변환을 통한 활용

y 가 $N(0,1)$ 이라면 $\sigma y + u$ 는 $N(u/\sigma^2)$ 를 따르게 된다.

Basic Sampling Algorithms (Rejection sampling)

<CDF 역함수 방법의 한계>

- CDF계산 적분과정의 어려움
- 역함수 구하는 것이 단순한 분포만 가능

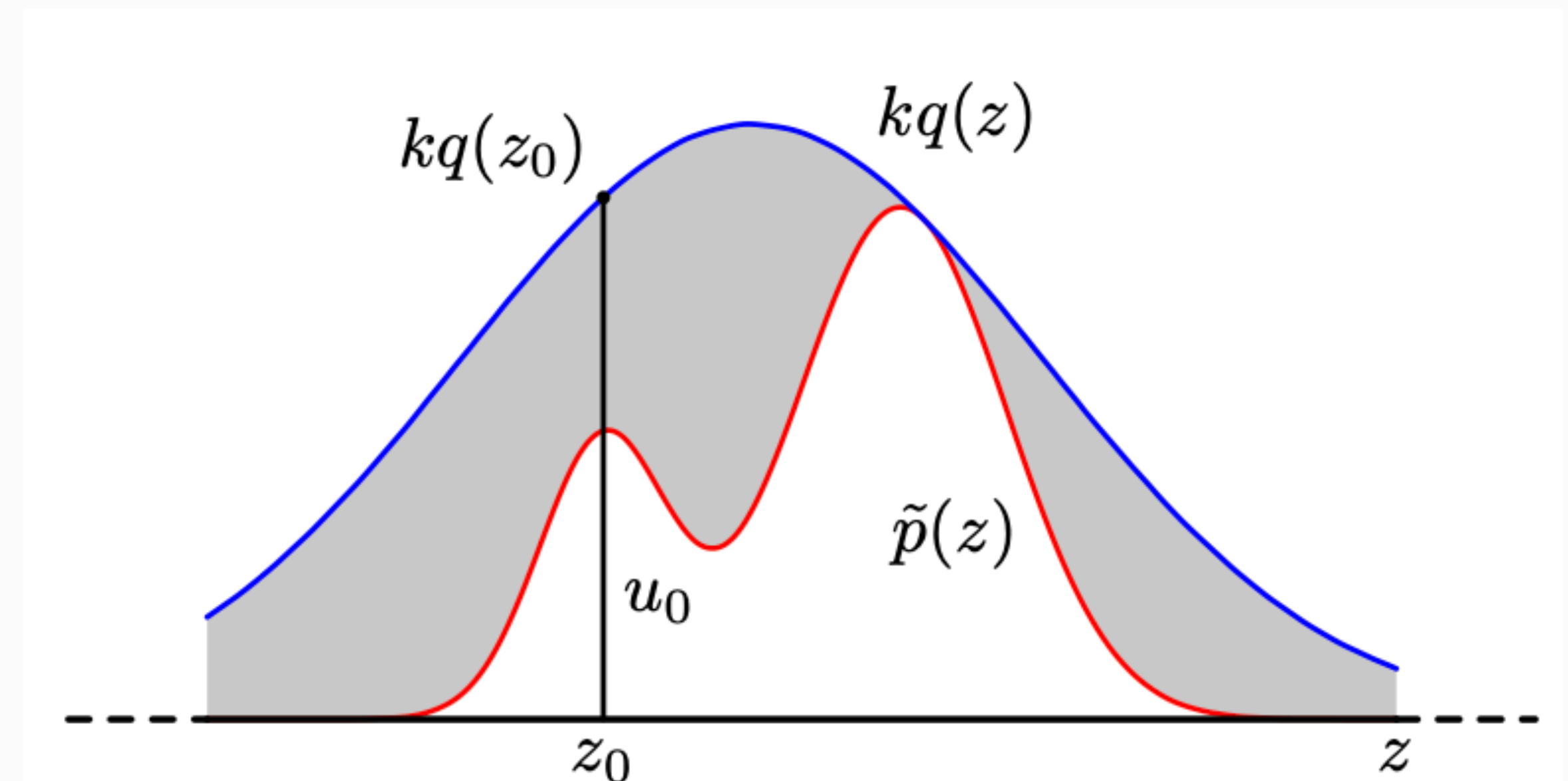
<Rejection sampling>

- 1) 정규화상수를 제외한 $p(z)$ 계산 가능

$$p(z) = \frac{1}{Z_p} \tilde{p}(z)$$

- 2) 표집이 가능한 단순한 제안 분포 $q(z)$ 설정

- 3) $kq(z)$ 을 통해 accept/rejection 을 결정



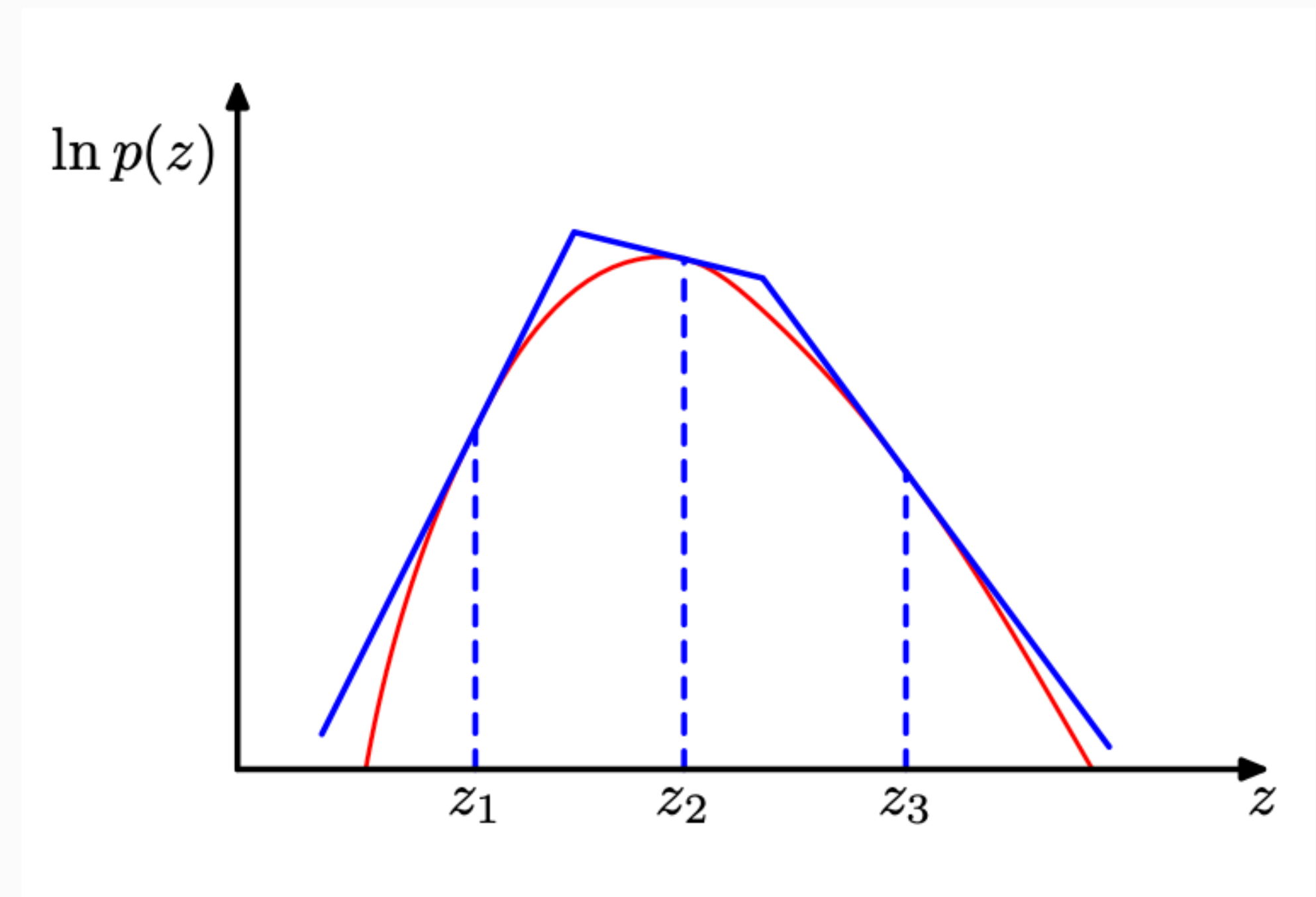
Basic Sampling Algorithms (Rejection sampling)

<적절한 제안분포 $q(z)$ 의 찾기의 어려움>

- $p(z)$ 가 로그오목일 경우, 쉬운방법이 존재

$$q(z) = k_i \lambda_i \exp \{-\lambda_i(z - z_{i-1})\} \quad z_{i-1} < z \leq z_i.$$

- 거부된 표본은 격자점의 집합에 추가한다.
- 점점 $p(z)$ 에 가까워지고 거부확률도 줄어들게 될것이다.
- 로그 오목이 아닐 경우,
- adaptive rejection Metropolis sampling 의 사용



Basic Sampling Algorithms (Adaptive rejection sampling)

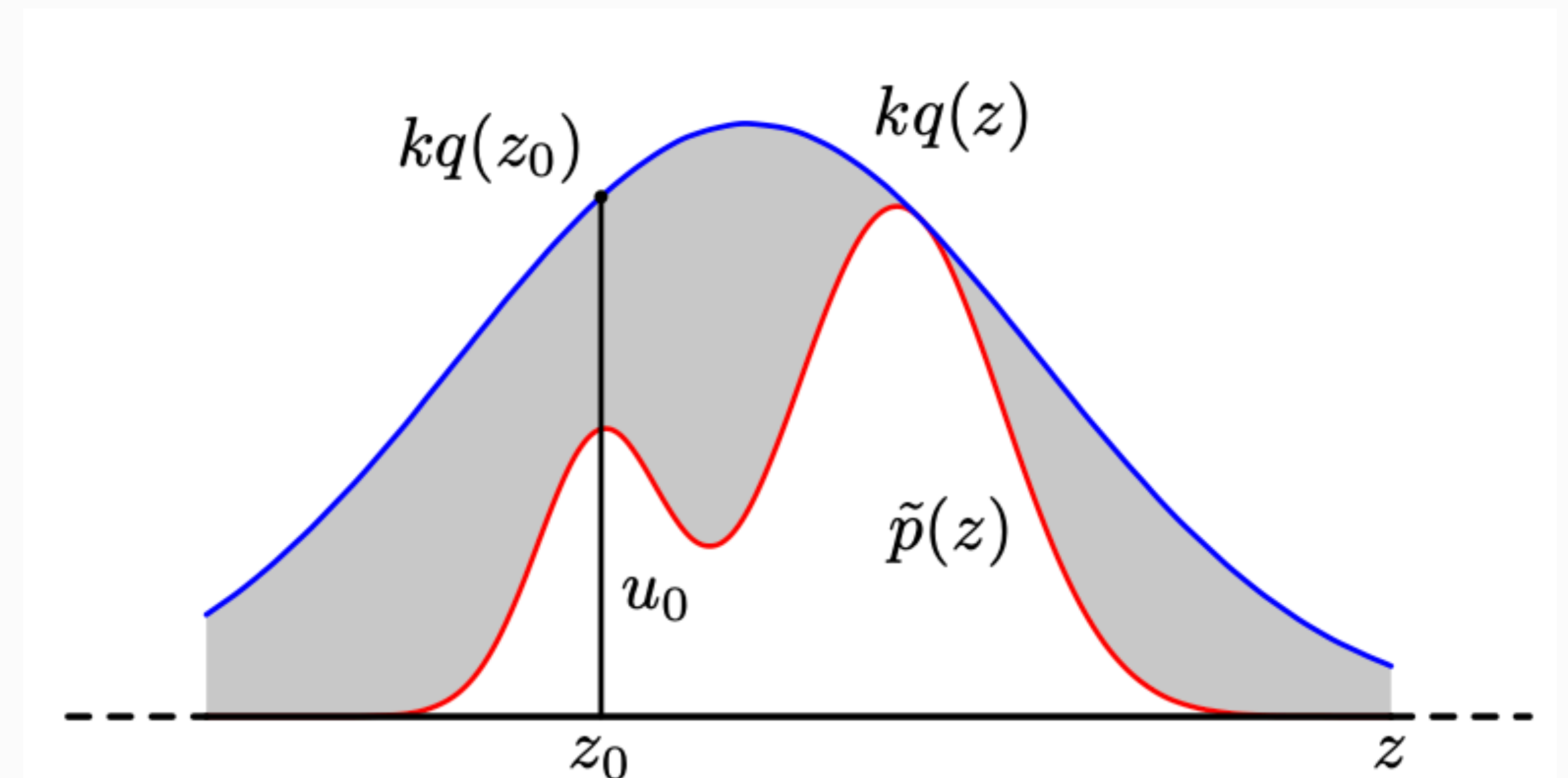
<Rejection Sampling>

- 1) 제안분포 $q(z)$ 로부터 샘플 z_0 생성
- 2) $[0, kq(z)]$ 에서 균일난수 u_0 생성
- 3) u 가 검정영역에 있지않으면 Accept
 u 가 검정영역에 있으면 Reject

<거부율의 최소화>

- 최대한 k 를 작게해야한다.
- 제안분포와 목표분포와 유사하게 설정해야한다.

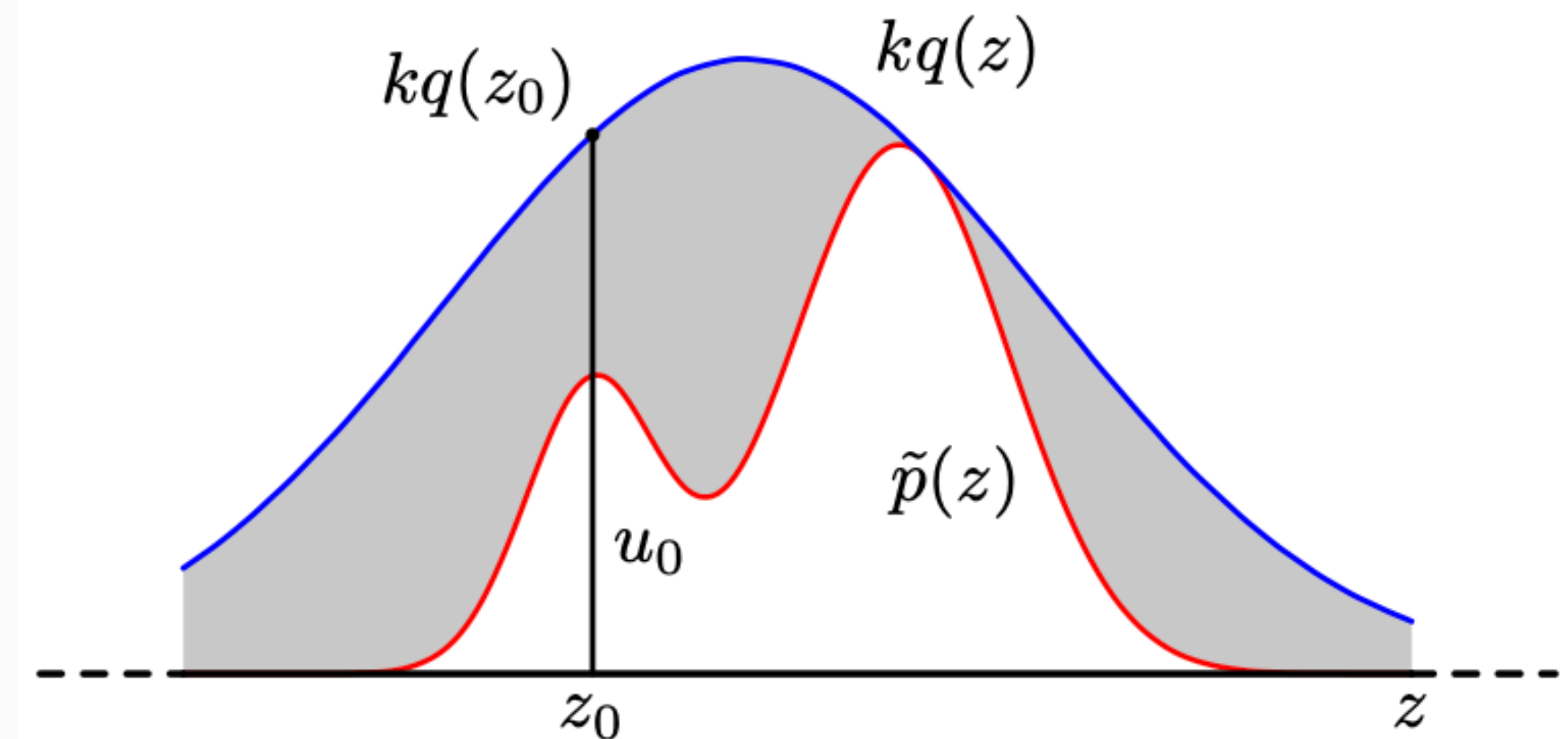
$$\begin{aligned} p(\text{accept}) &= \int \{\tilde{p}(z)/kq(z)\} q(z) dz \\ &= \frac{1}{k} \int \tilde{p}(z) dz. \end{aligned}$$



Basic Sampling Algorithms (Importance sampling)

<Rejection sampling 의 한계>

- 적절한 상수 k 를 결정의 어려움
- 너무 커지면 낮은 승인률 실용성x
- 너무 작으면 원분포의 경계를 보장 못함
- **Importance sampling resamplings**는 k 를 필요로 하지 않는다.



Basic Sampling Algorithms (Importance sampling)

<Importance sampling>

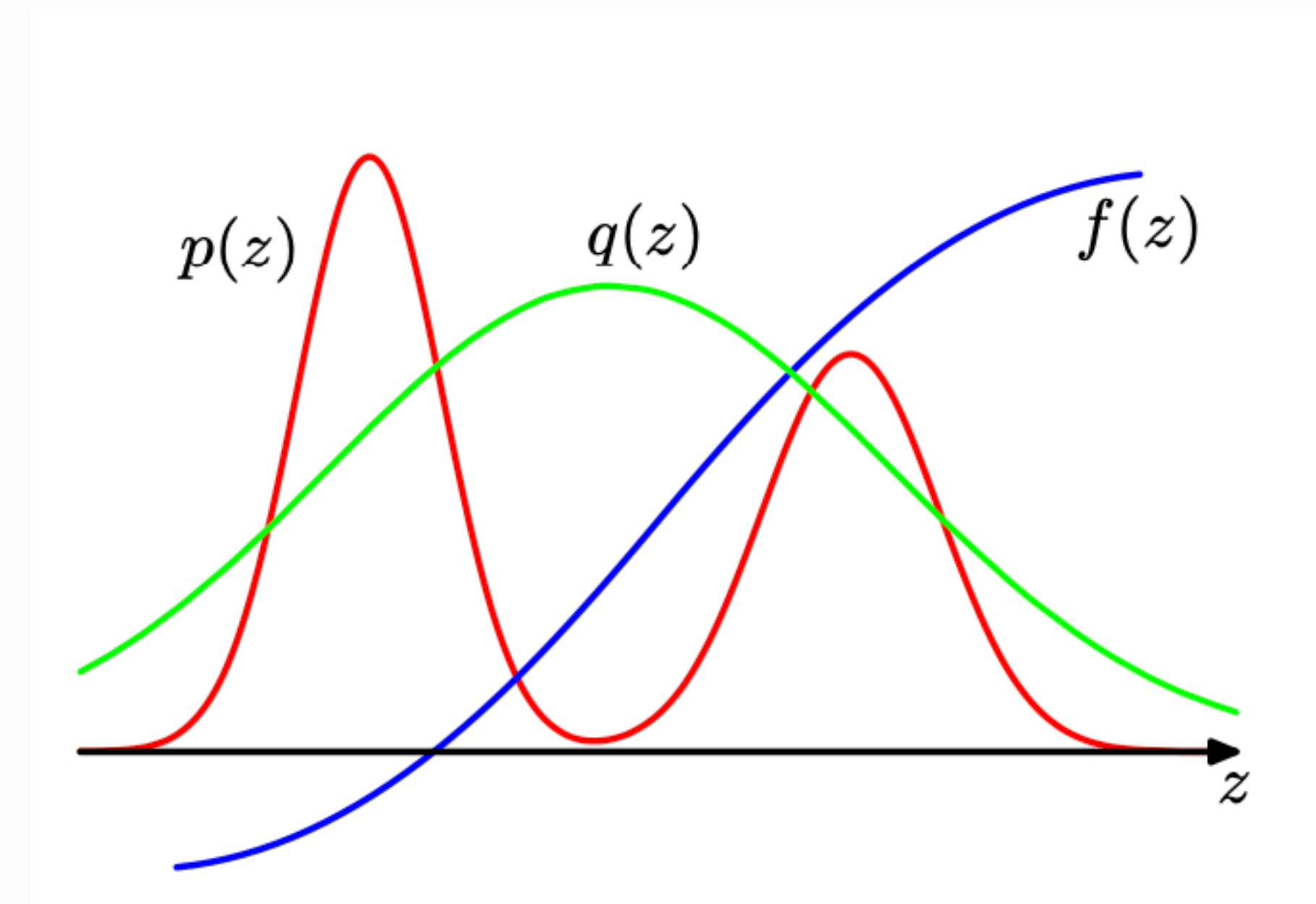
목표: Y 에서 랜덤한 수 생성

-> 목표: $E[f]$ 를 직접 근사

$$\mathbb{E}[f] \simeq \sum_{l=1}^L p(\mathbf{z}^{(l)}) f(\mathbf{z}^{(l)}).$$

문제: 차원수가 커짐에 따라 합산항의 숫자가
기하급수적으로 늘어난다

해결: $p(z)$ 가 큰구역에서 가중치를 뒤서 표본선택
—> Importance sampling



Basic Sampling Algorithms (Importance sampling)

<Importance sampling>

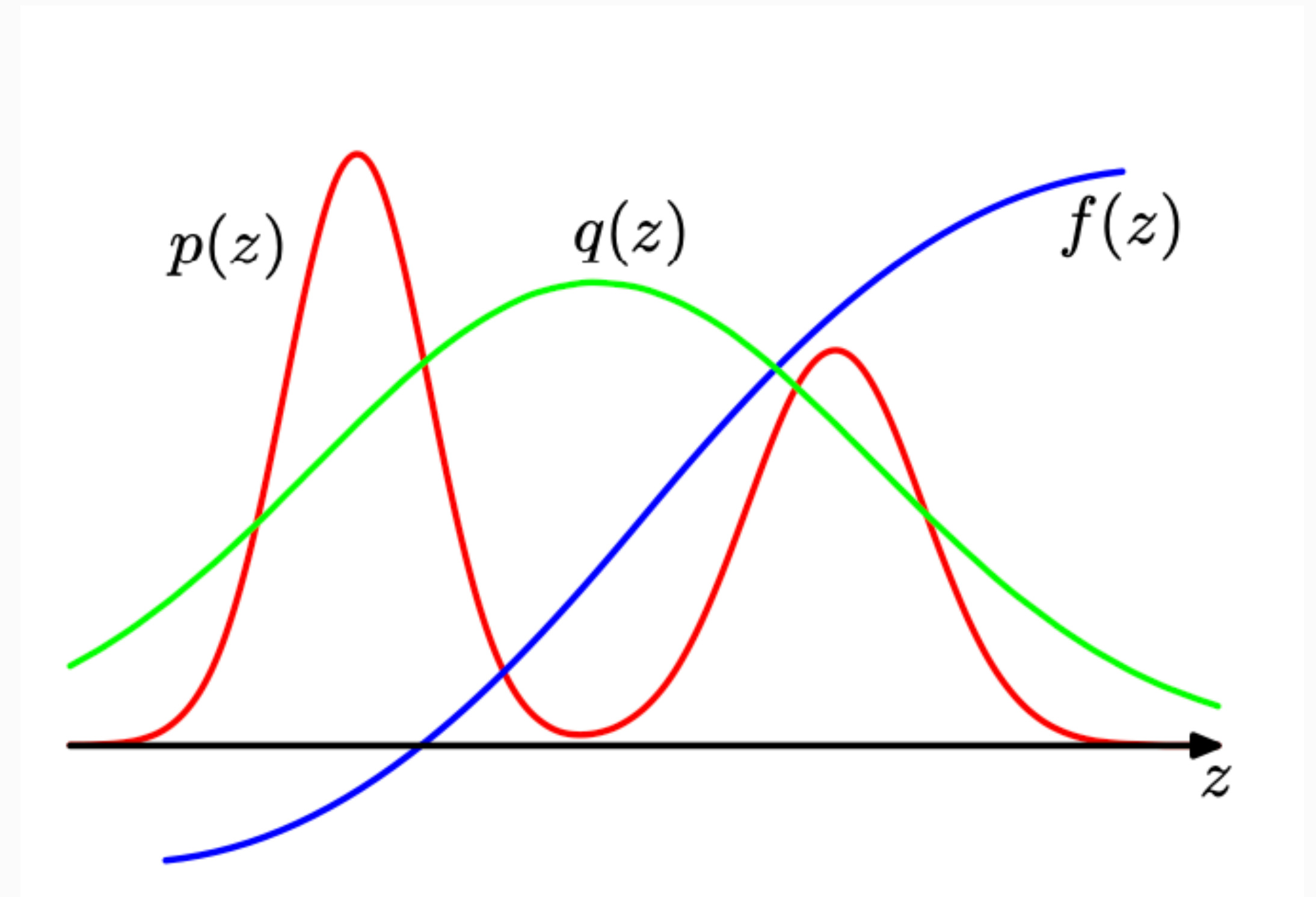
- 1) 제안분포 $q(z)$ 에서 표집
- 2) 중요도 가중치에 따라 표집

$$r_l = p(\mathbf{z}^{(l)}) / q(\mathbf{z}^{(l)})$$

- 3) 기댓값의 근사

$$\begin{aligned}\mathbb{E}[f] &= \int f(\mathbf{z}) p(\mathbf{z}) d\mathbf{z} \\ &= \int f(\mathbf{z}) \frac{p(\mathbf{z})}{q(\mathbf{z})} q(\mathbf{z}) d\mathbf{z} \\ &\simeq \frac{1}{L} \sum_{l=1}^L \frac{p(\mathbf{z}^{(l)})}{q(\mathbf{z}^{(l)})} f(\mathbf{z}^{(l)}).\end{aligned}$$

- 우리는 정규화상수를 모르기때문에 위 식을 계산할 수 없다.



Basic Sampling Algorithms (Importance sampling)

<Importance sampling>

1) 제안분포 $q(\mathbf{z})$ 에서 표집

2) 중요도 가중치에 따라 표집

$$r_l = p(\mathbf{z}^{(l)})/q(\mathbf{z}^{(l)})$$

3) 기댓값의 근사 (계산가능하게)

$$\begin{aligned}\mathbb{E}[f] &= \int f(\mathbf{z})p(\mathbf{z}) d\mathbf{z} \\ &= \frac{Z_q}{Z_p} \int f(\mathbf{z}) \frac{\tilde{p}(\mathbf{z})}{\tilde{q}(\mathbf{z})} q(\mathbf{z}) d\mathbf{z} \\ &\simeq \frac{Z_q}{Z_p} \frac{1}{L} \sum_{l=1}^L \tilde{r}_l f(\mathbf{z}^{(l)}).\end{aligned}$$

4) 정규화상수 계산

$$\begin{aligned}\frac{Z_p}{Z_q} &= \frac{1}{Z_q} \int \tilde{p}(\mathbf{z}) d\mathbf{z} = \int \frac{\tilde{p}(\mathbf{z})}{\tilde{q}(\mathbf{z})} q(\mathbf{z}) d\mathbf{z} \\ &\simeq \frac{1}{L} \sum_{l=1}^L \tilde{r}_l\end{aligned}$$

5) 기댓값의 근사 (간단하게정리)

$$\mathbb{E}[f] \simeq \sum_{l=1}^L w_l f(\mathbf{z}^{(l)})$$

$$w_l = \frac{\tilde{r}_l}{\sum_m \tilde{r}_m} = \frac{\tilde{p}(\mathbf{z}^{(l)})/q(\mathbf{z}^{(l)})}{\sum_m \tilde{p}(\mathbf{z}^{(m)})/q(\mathbf{z}^{(m)})}.$$

Basic Sampling Algorithms (Importance sampling)

<Importance sampling resamplings>

- 첫번째, $q(z)$ 로 부터 L 개의 표본을 추출한다. $\mathbf{z}^{(1)}, \dots, \mathbf{z}^{(L)}$
- 두번째, 가중치 w_1, \dots, w_L 을 구성한다
- 세번째, L 개의 표본으로 부터 resampling한다. 이때 가중치에 따라 추출한다.
- 이 결과로 얻게 된 L 개의 표본들은 $p(z)$ 에 대해서 근사적으로만 분포된다.
- 하지만, $\lim L$ 이 무한대로 갈때 올바르게 된다.

$$w_l = \frac{\tilde{r}_l}{\sum_m \tilde{r}_m} = \frac{\tilde{p}(\mathbf{z}^{(l)})/q(\mathbf{z}^{(l)})}{\sum_m \tilde{p}(\mathbf{z}^{(m)})/q(\mathbf{z}^{(m)})}.$$

$$\begin{aligned} p(z \leq a) &= \sum_{l: z^{(l)} \leq a} w_l \\ &= \frac{\sum_l I(z^{(l)} \leq a) \tilde{p}(z^{(l)})/q(z^{(l)})}{\sum_l \tilde{p}(z^{(l)})/q(z^{(l)})} \end{aligned}$$

$$\begin{aligned} p(z \leq a) &= \frac{\int I(z \leq a) \{\tilde{p}(z)/q(z)\} q(z) dz}{\int \{\tilde{p}(z)/q(z)\} q(z) dz} \\ &= \frac{\int I(z \leq a) \tilde{p}(z) dz}{\int \tilde{p}(z) dz} \\ &= \int I(z \leq a) p(z) dz \end{aligned}$$

$p(z)$ 의 누적분포가
됨을 볼 수 있다.
이때 $p(z)$ 의 정규화과정은
필요없다.

2

MCMC

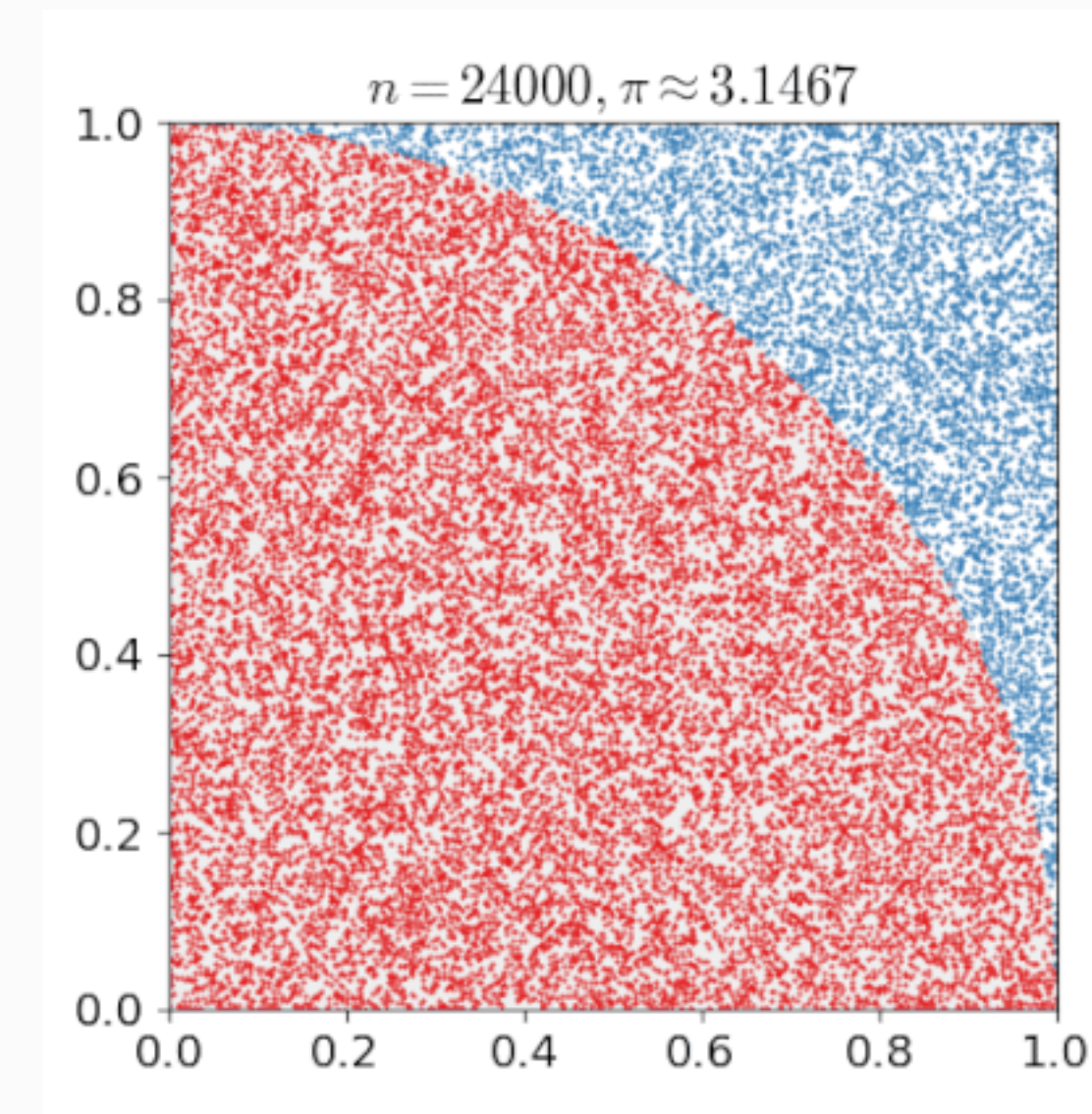
MCMC Markov Chain Monte Carlo

<MCMC>

- 고차원 공간상에서 심각한 한계를 가지고 있는
- rejection/importance sampling의 해결책
- MCMC= Monte Carlo + Markov

<Montecarlo 알고리즘>

수학적인 결과를 얻기 위해 반복적으로 무작위 샘플링의 방법을 이용하는 넓은 범위의 컴퓨터 알고리즘이다.



MCMC Markov Chain Monte Carlo

< Markov chain >

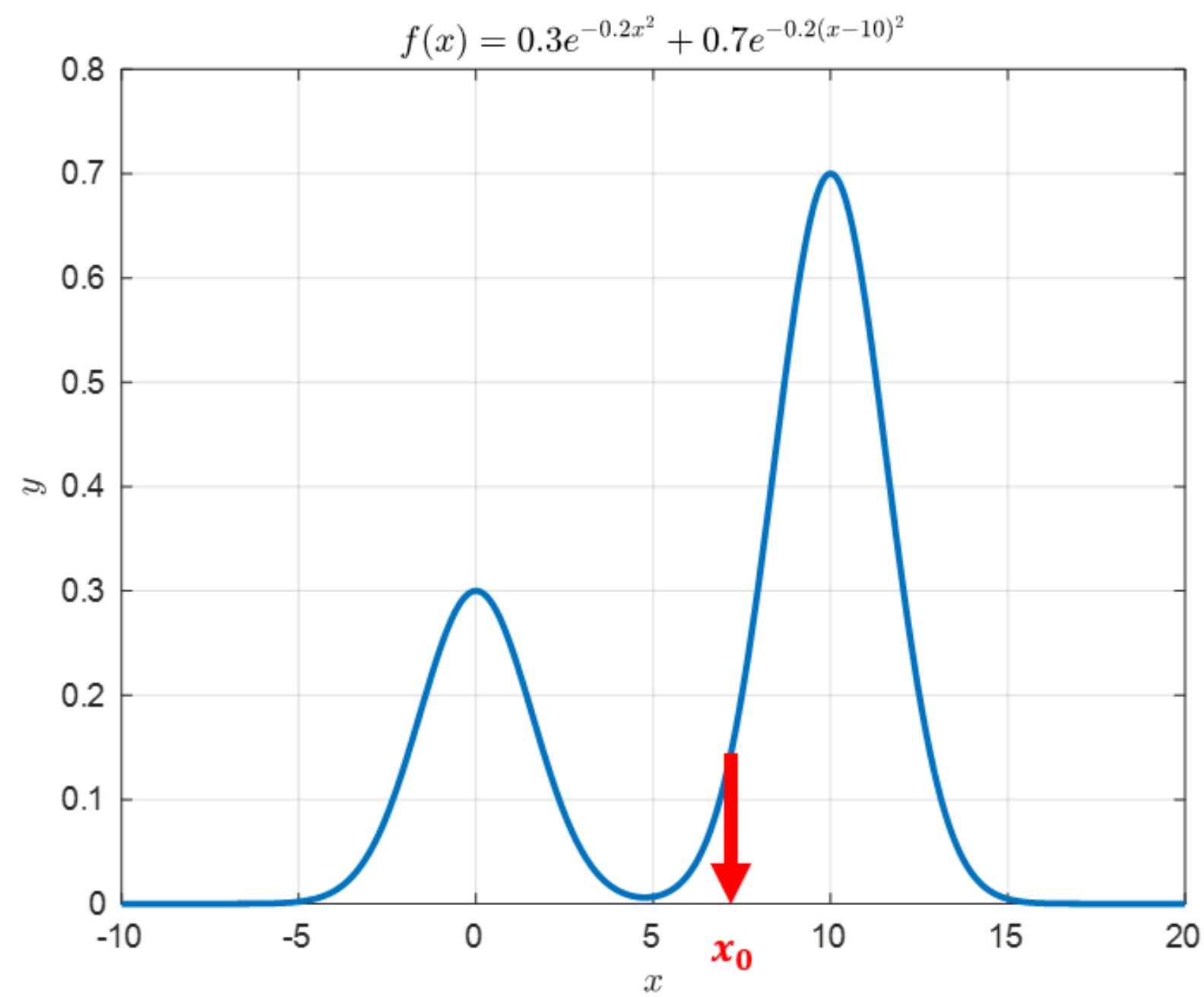
- 마르코프 체인의 정의란 ‘마르코프 성질’을 가진 ‘이산 확률과정’
- 여기서 마르코프 성질은 ‘과거와 현재 상태가 주어졌을 때, 미래 상태의 조건부 확률분포가 과거 상태에 영향을 받지 않고 독립적으로 현재 상태로만 결정되는 것을 의미’ (조건부 독립성이 성립)
- 예를 들어 오늘의 날씨가 맑다면 내일의 날씨는 맑을지 비가 내릴지를 확률적으로 표현할 수 있다.

$$p(\mathbf{z}^{(m+1)} | \mathbf{z}^{(1)}, \dots, \mathbf{z}^{(m)}) = p(\mathbf{z}^{(m+1)} | \mathbf{z}^{(m)}).$$

<Metropolis algorithms >

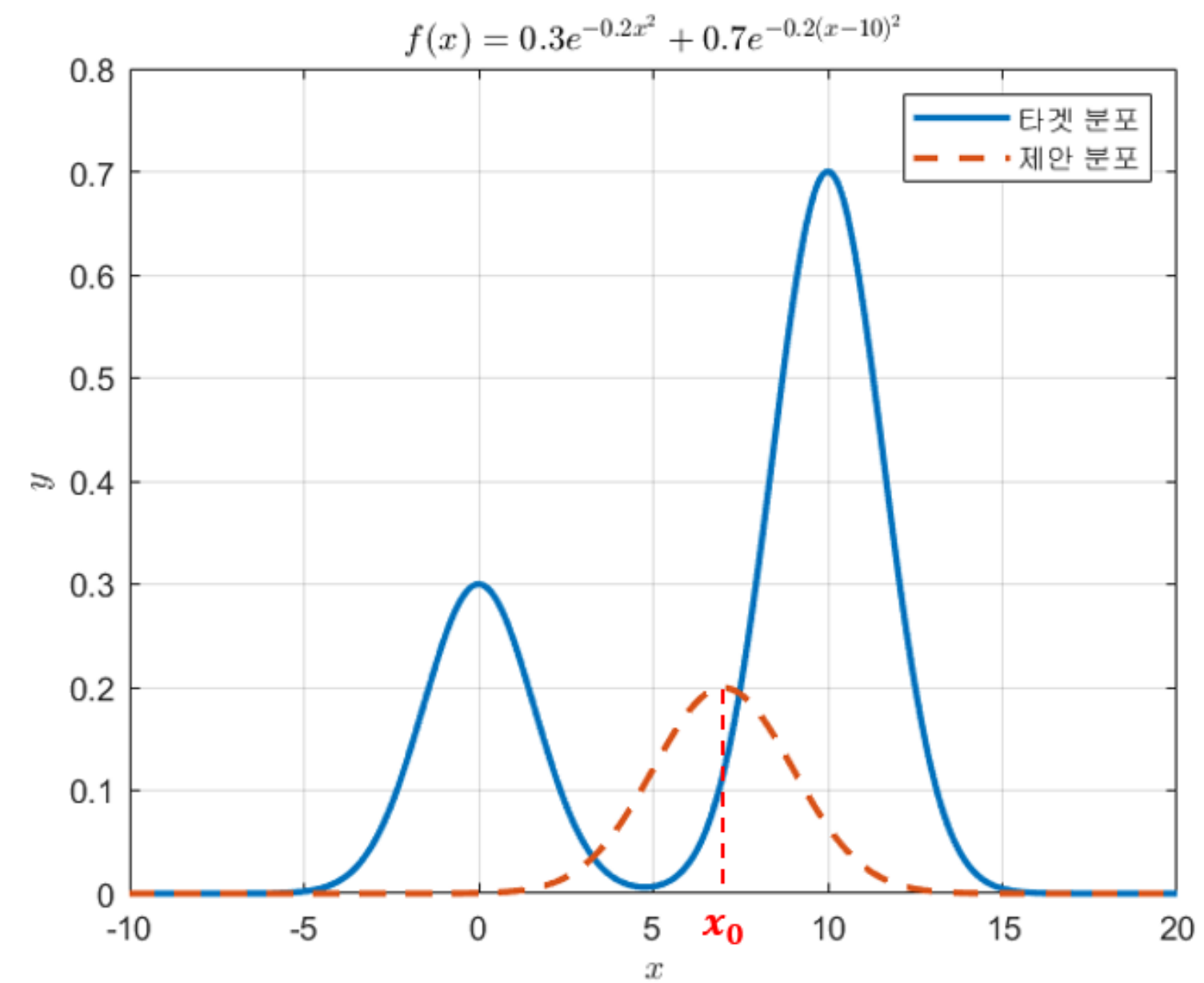
목표: 복잡한 target 분포 $f(z)$ 에서의 샘플생성

1) Random initialization : 첫샘플을 아무거나 선택



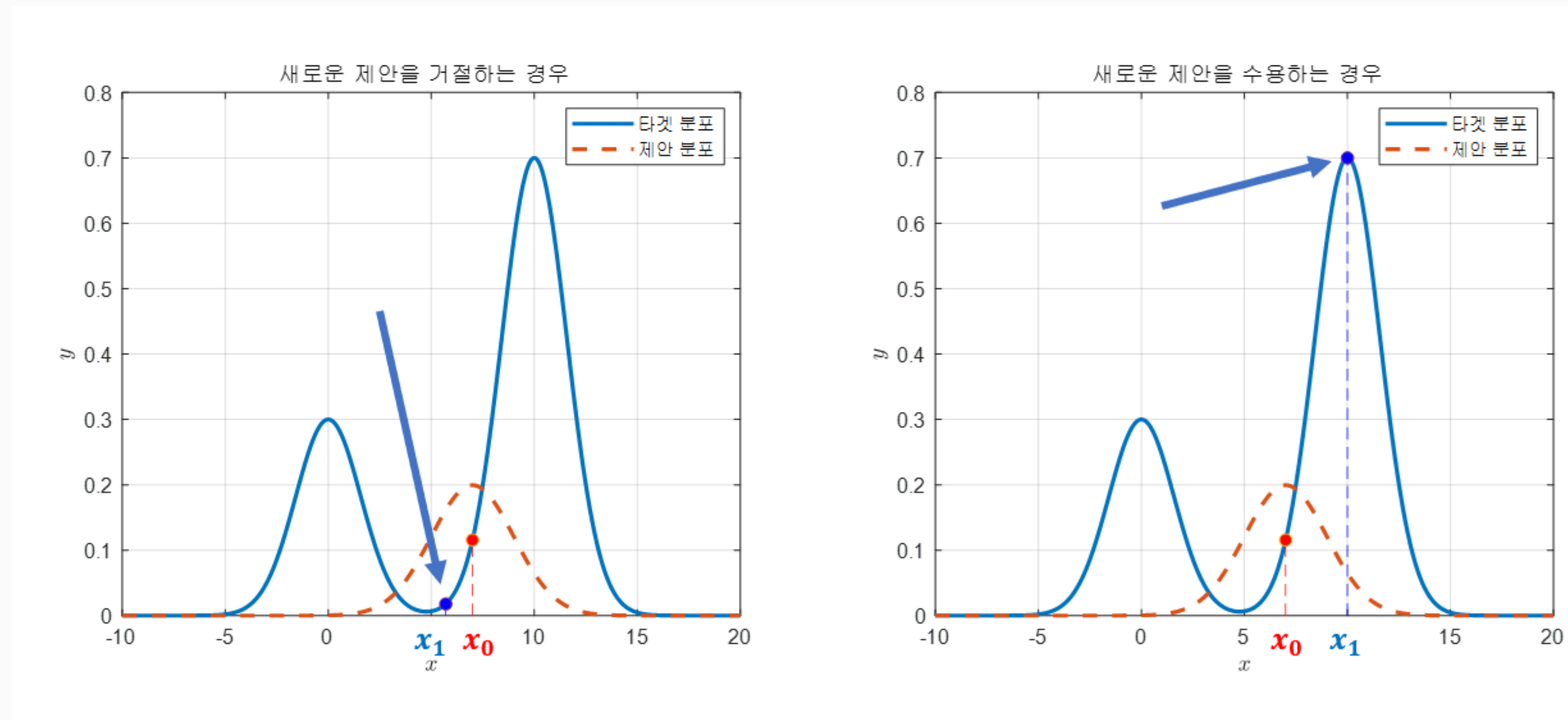
2) 제안분포로부터 다음 샘플을 생성 (symmetric한 확률분포)

- $z^{(t)}$ 에 종속적인 제안분포 $q(z | z(\tau))$



<Metropolis algorithms >

3) 제안분포로부터 생성된 샘플 Accept or Reject



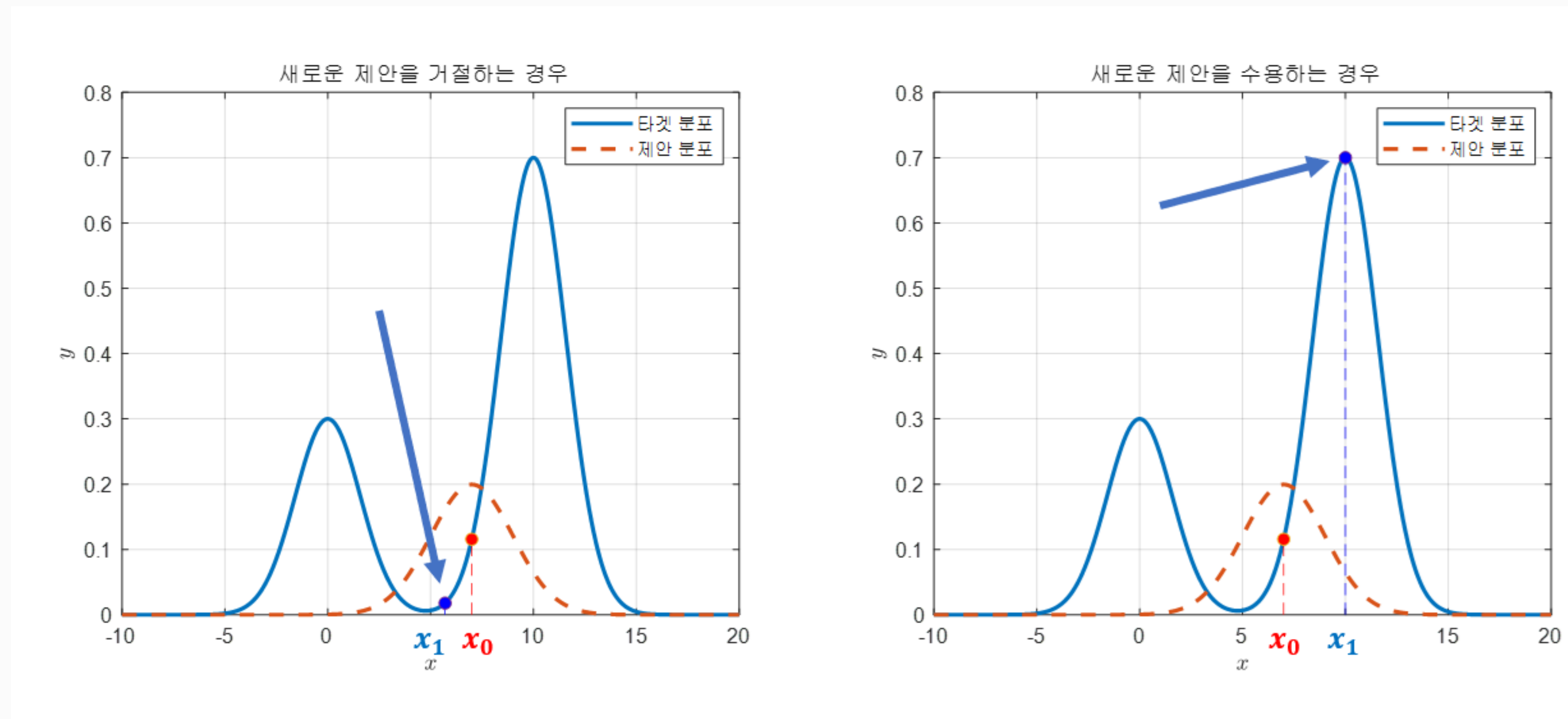
Accept 기준

$$\frac{f(x_1)}{f(x_0)} > 1$$

- 여기서 accept 하지 못했다면 바로 reject하는 것이 아니라 한번 더 검사를 하게된다.
- why?

<Metropolis algorithms >

3) 제안분포로 부터 생성된 샘플 Accept or Reject



새로운 Accept 기준

$$\frac{f(x_1)}{f(x_0)} > u$$

u 는 $[0,1]$ 유니폼난수

<Metropolis algorithms >

단순한 Random walk 알고리즘 예시

$$\begin{aligned}p(z^{(\tau+1)} = z^{(\tau)}) &= 0.5 \\p(z^{(\tau+1)} = z^{(\tau)} + 1) &= 0.25 \\p(z^{(\tau+1)} = z^{(\tau)} - 1) &= 0.25\end{aligned}$$

- $z^{(0)}=0$ 이면 대칭성에 따라서 $z^{(t)}$ 에서의 기댓값 $E[z^{(t)}]=0$, $E[z^{(t)^2}]=t/2$ 일 것이다.
- 제곱근 종속성의 발생: t 단계 지난 후에 임의의 보행은 평균적으로 t 의 제곱근에 해당하는 만큼의 거리를 이동하게 될 것이다.
- 이로 부터 Random walk 알고리즘이 매우 비효율적이고 MCMC는 이러한 행태를 피하고자 한다.

<Markov chains >

<1차 마르코프연쇄 : 조건부 독립성이 성립>

$$p(\mathbf{z}^{(m+1)} | \mathbf{z}^{(1)}, \dots, \mathbf{z}^{(m)}) = p(\mathbf{z}^{(m+1)} | \mathbf{z}^{(m)}).$$

<Transition probabilities (전이확률)>

$$p(\mathbf{z}^{(m+1)}) = \sum_{\mathbf{z}^{(m)}} p(\mathbf{z}^{(m+1)} | \mathbf{z}^{(m)}) p(\mathbf{z}^{(m)}).$$

<homogeneous(동질적)>

모든 m 에 대해 전이확률이 동일한 마르코프 연쇄를 homogeneous(동질적)이라고 한다.

<Markov chains >

<invariant(불변적)>

마르코프 연쇄에서의 각 단계가 분포를 변하지 않은 채로 내버려 둘 경우
그 분포를 해당 마르코프 연쇄에 대해서 invariant(불변적) stationary(정류적)이라고 한다.

따라서 전이확률이 동질적일때, 다음을 만족하면 불변적이다.

$$p^*(\mathbf{z}) = \sum_{\mathbf{z}'} T(\mathbf{z}', \mathbf{z}) p^*(\mathbf{z}').$$

<detailed balance(세부균형) 과 reversible(가역적)>

전이확률이 다음의 detailed balance을 만족하는건 원분포가 불변적이기 위한 충분조건이다.

$$p^*(\mathbf{z}) T(\mathbf{z}, \mathbf{z}') = p^*(\mathbf{z}') T(\mathbf{z}', \mathbf{z})$$

세부 균형을 만족하는 마르코프 연쇄를 reversible이라고 한다.

<Markov chains >

<ergdicity (에르고딕성)>

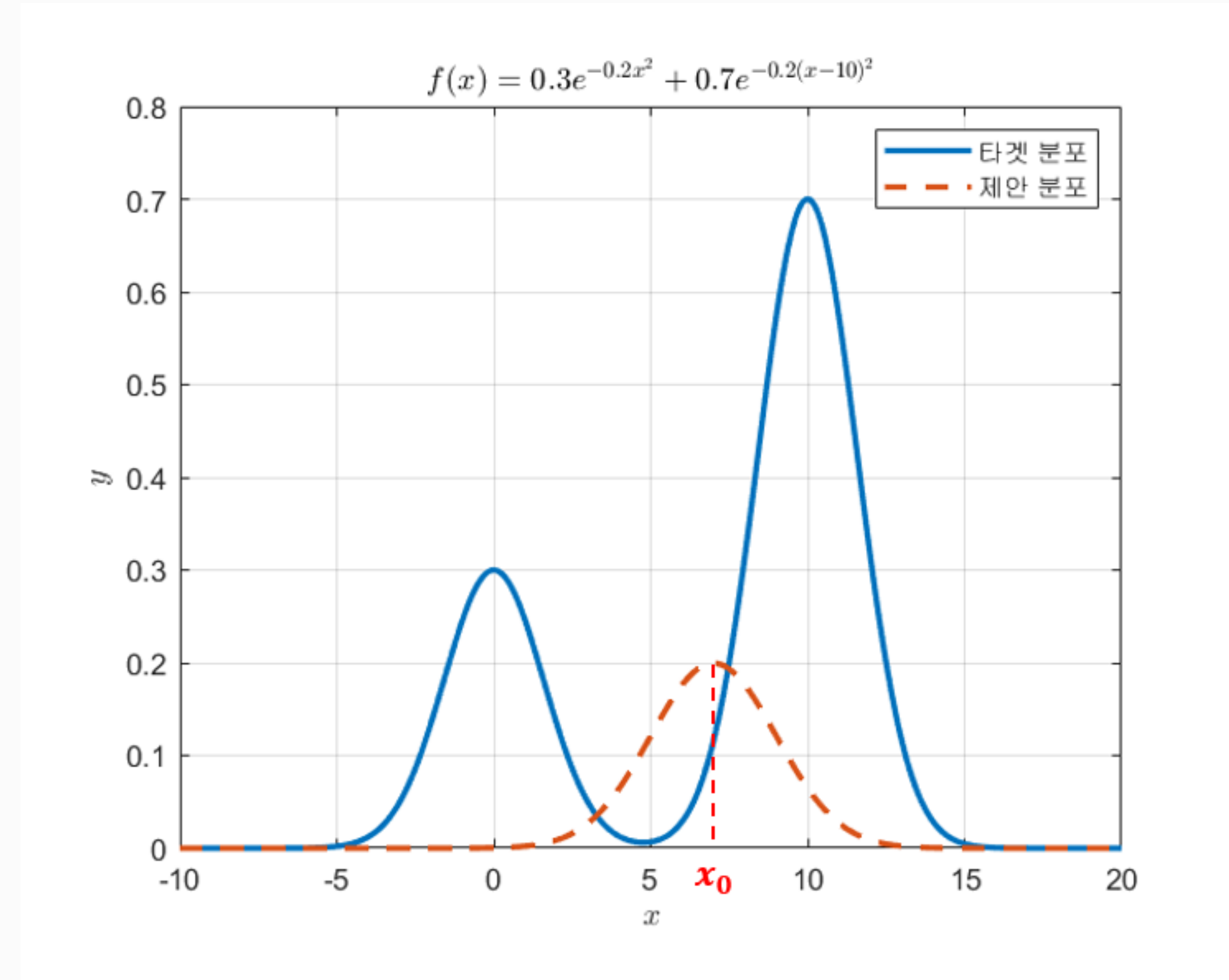
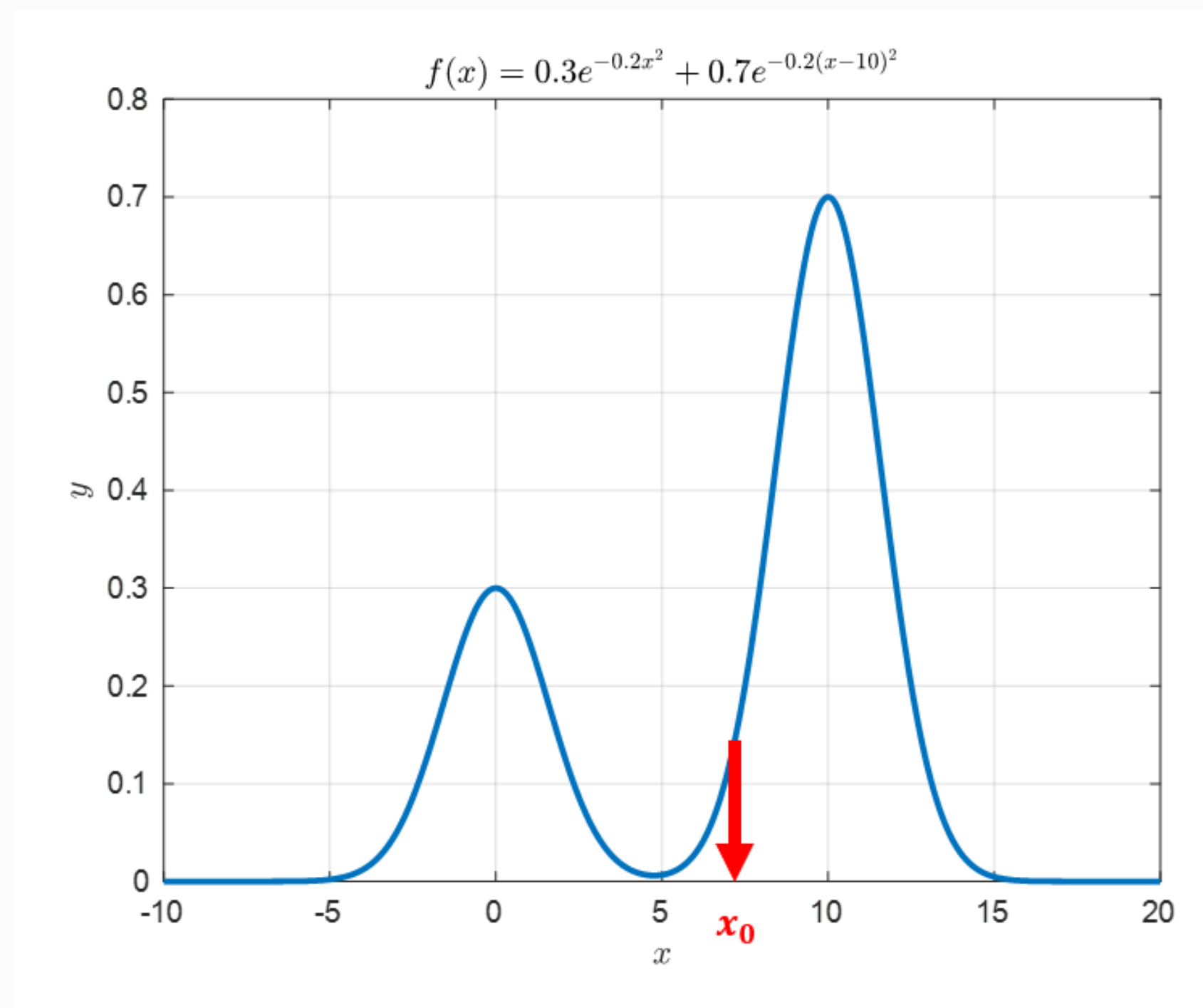
- m 이 무한대로 갈때 어떤 초기분포 $p(z^{(0)})$ 을 선택하든 상관 없이 $p(z^{(m)})$ 이 해당 불변분포 $p^*(z)$ 로 수렴하면 ergdicity를 가진다
- 이 경우 불변분포를 equilibrium distribution 평형분포 라고 하고 에르고딕성을 지니는 마르코프 연쇄 하나는 단 하나의 평형 분포만을 가진다.

<Metropolis-Hastings algorithms >

- 복잡한 target 분포 $f(z)$ 가정

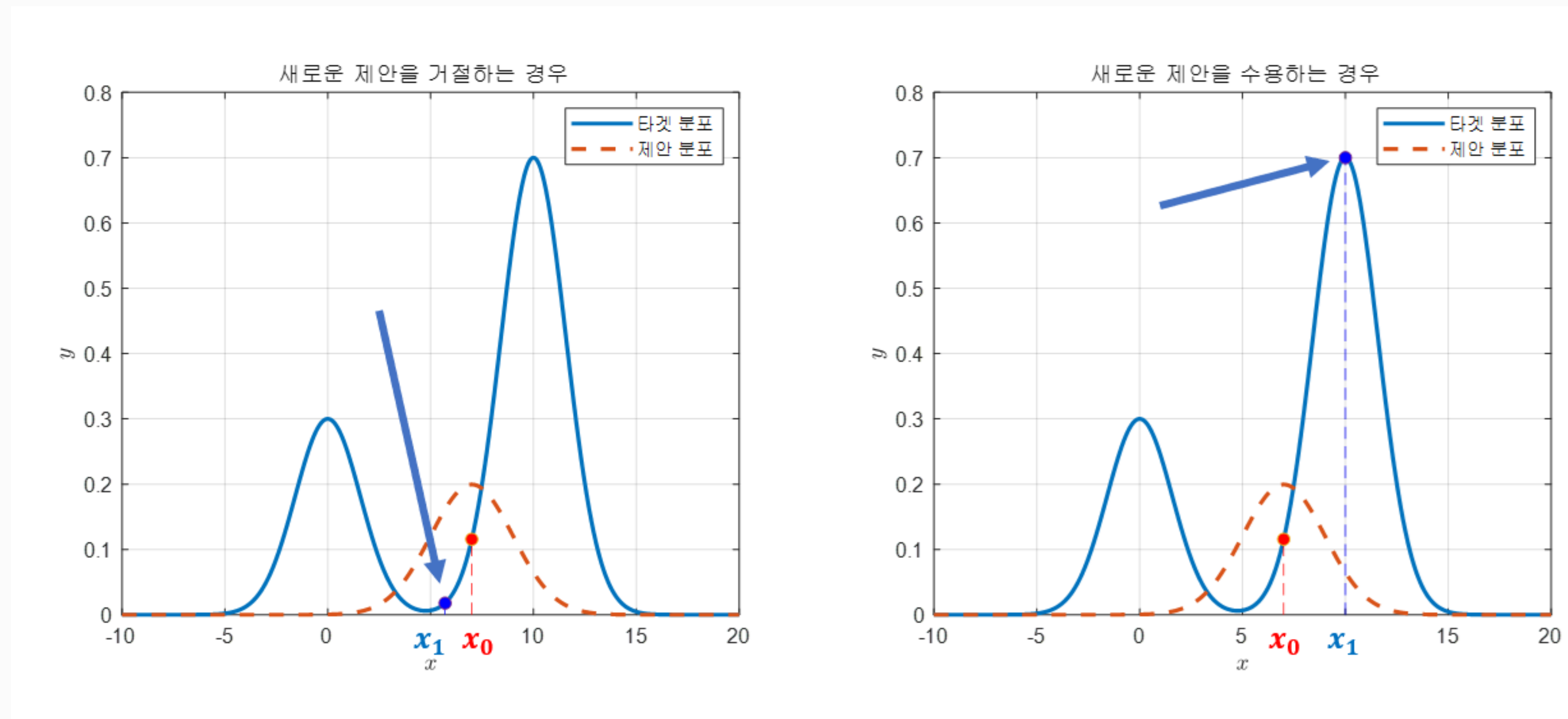
1) Random initialization : 첫샘플을 랜덤하게 선택

2) 제안분포로부터 다음 샘플을 생성 (**symmetric한 확률분포**)



<Metropolis-Hastings algorithms >

3) 제안분포로 부터 생성된 샘플 Accept or Reject



원래 Accept 기준

$$\frac{f(x_1)}{f(x_0)} > 1$$

새로운 Accept 기준

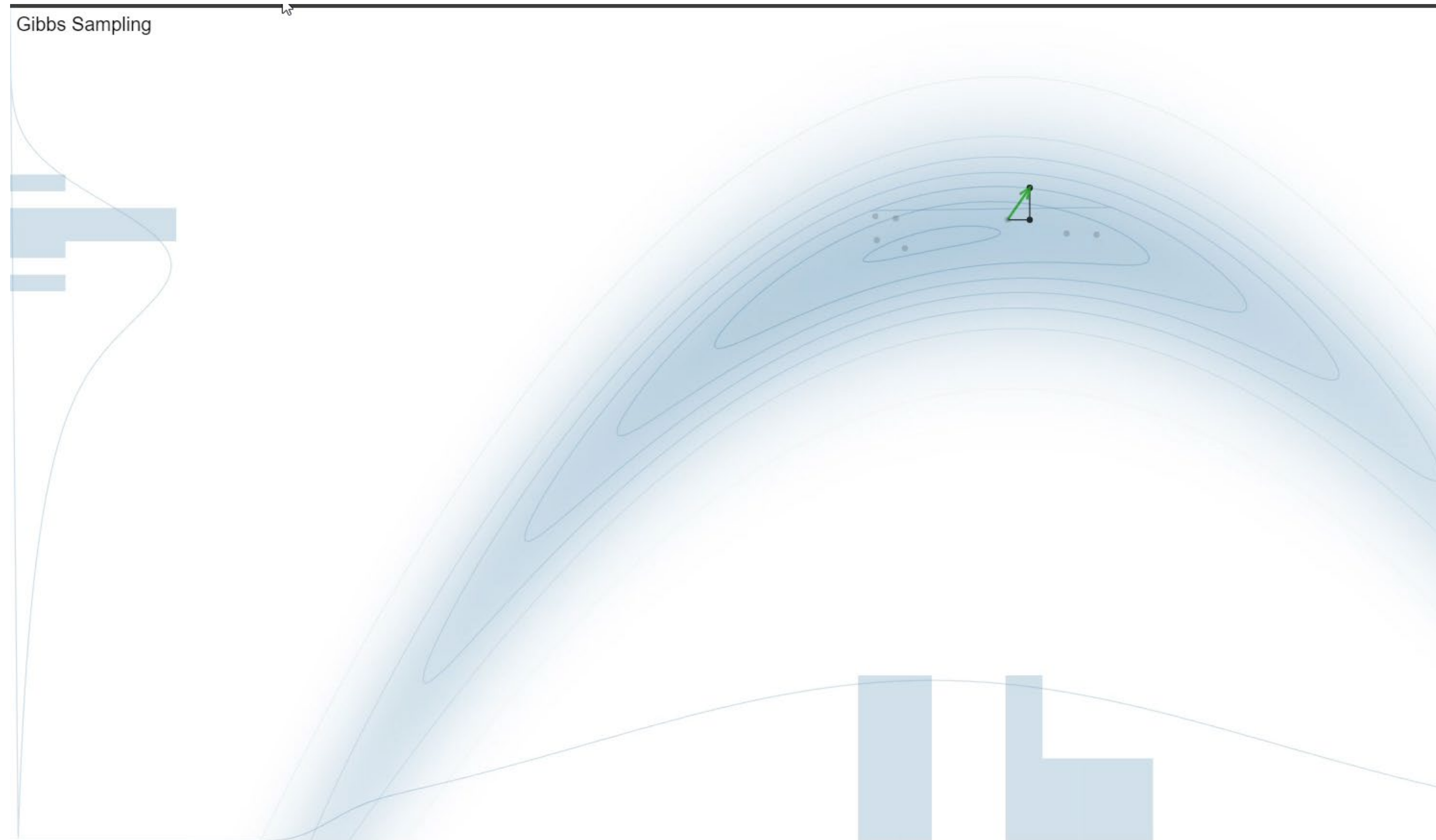
$$\frac{f(x_1)/g(x_1|x_0)}{f(x_0)/g(x_0|x_1)} > 1$$

11.3

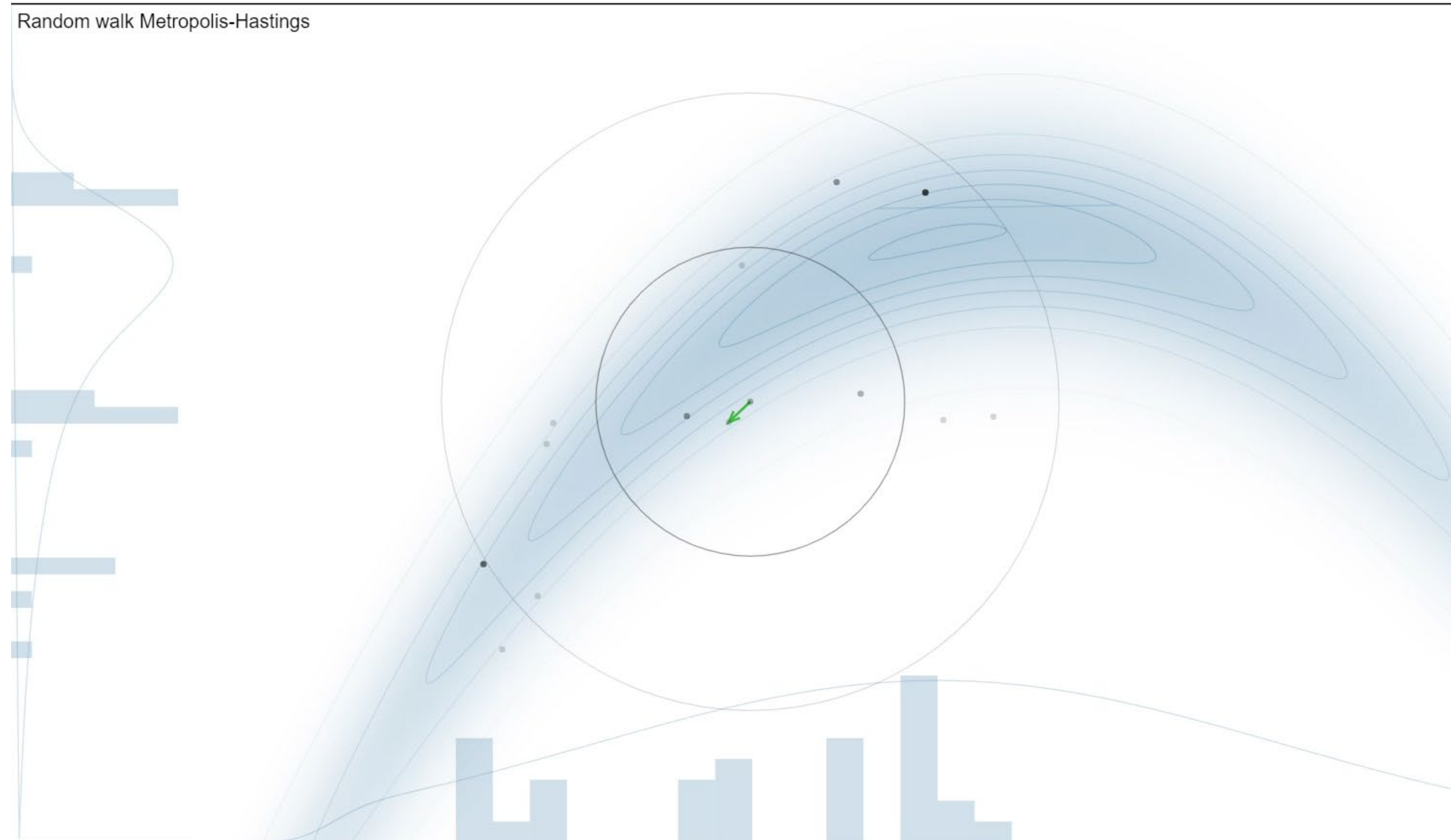
Gibbs Sampling

Special case of the Metropolis-Hastings algorithm

Gibbs Sampling - 직관적 예시 ; 시뮬레이션



Gibbs Sampling - 직관적 예시 vs Random walk MH



Gibbs Sampling - 그냥 예시 (i=3)

- 변수 3개에 대한 분포 $p(z_1, z_2, z_3)$
- 알고리즘의 각 단계에서 선택한 값들 $z_1^{(\tau)}, z_2^{(\tau)}, z_3^{(\tau)}$
- 다음 단계의 z_1, z_2, z_3 를 선택하게 하는 조건부 분포

$$p(z_1 | z_2^{(\tau)}, z_3^{(\tau)})$$

$$p(z_2 | z_1^{(\tau+1)}, z_3^{(\tau)})$$

$$p(z_3 | z_1^{(\tau+1)}, z_2^{(\tau+1)})$$

- 3개의 변수들을 순서대로 돌아가면서 위의 과정을 반복하게 된다.

Gibbs Sampling - 일반화

Gibbs Sampling

1. Initialize $\{z_i : i = 1, \dots, M\}$
2. For $\tau = 1, \dots, T$:
 - Sample $z_1^{(\tau+1)} \sim p(z_1 | z_2^{(\tau)}, z_3^{(\tau)}, \dots, z_M^{(\tau)})$.
 - Sample $z_2^{(\tau+1)} \sim p(z_2 | z_1^{(\tau+1)}, z_3^{(\tau)}, \dots, z_M^{(\tau)})$.
 - \vdots
 - Sample $z_j^{(\tau+1)} \sim p(z_j | z_1^{(\tau+1)}, \dots, z_{j-1}^{(\tau+1)}, z_{j+1}^{(\tau)}, \dots, z_M^{(\tau)})$.
 - \vdots
 - Sample $z_M^{(\tau+1)} \sim p(z_M | z_1^{(\tau+1)}, z_2^{(\tau+1)}, \dots, z_{M-1}^{(\tau+1)})$.

Gibbs Sampling

- 샘플링을 위한 제안 분포가 필요하지 않다

⇔ 모수에 대한 사후 조건부 분포가 잘 알려진 형태로 나타날 경우, 사후 조건부 분포에서의 샘플링을 통해 사후 분포를 추정할 수 있게 된다.

표본 X 가 $N(\mu, \sigma^2)$ 분포에서 생성된 n 개의 샘플이라고 가정하자.

unknown parameter.

let μ 의 사전 분포 : $N(0, \tau^2)$

⇒ μ 의 사후분포 $p(\mu, \sigma^2 | X) \Rightarrow$ 잘 알려진 분포 \times

σ^2 의 사전 분포 : $\text{Inv-Gamma}(\alpha, \beta)$

(계산 과정)

$$p(\mu, \sigma^2 | X) \propto p(X | \mu, \sigma^2) p(\mu) p(\sigma^2)$$

$$\propto \exp \left[-\frac{n}{2\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 - \frac{\mu^2}{2\tau^2} \right] \tau^{-(\frac{n}{2} + \alpha) - 1} \exp \left[-\frac{\beta}{\sigma^2} \right]$$

Gibbs Sampling

* 만약 μ, σ^2 각각의 조건부 분포를 살펴보자:

$$\begin{aligned} \textcircled{1} \quad p(\mu | \sigma^2, X) &\propto \exp \left[-\frac{n}{2\sigma^2} (\mu^2 - 2\mu \bar{X}) - \frac{\mu^2}{2\sigma^2} \right] \\ &\propto \exp \left[-\frac{1}{2} \left[\frac{n}{\sigma^2} + \frac{1}{\sigma^2} \right] \mu^2 - 2\mu \frac{n}{\sigma^2} \bar{X} \right] \\ &\propto \exp \left[-\frac{1}{2} \left(\frac{n}{\sigma^2} + \frac{1}{\sigma^2} \right) \left(\mu - \frac{\frac{n\bar{X}}{\sigma^2}}{\left(\frac{n}{\sigma^2} + \frac{1}{\sigma^2} \right)} \right)^2 \right] \\ \textcircled{2} \quad p(\sigma^2 | \mu, X) &\propto (\sigma^2)^{-(\frac{1}{2} + \alpha)} \exp \left[-\frac{1}{\sigma^2} \left[\frac{n}{2} \sum_{i=1}^n (x_i - \mu)^2 + \beta \right] \right] \end{aligned}$$

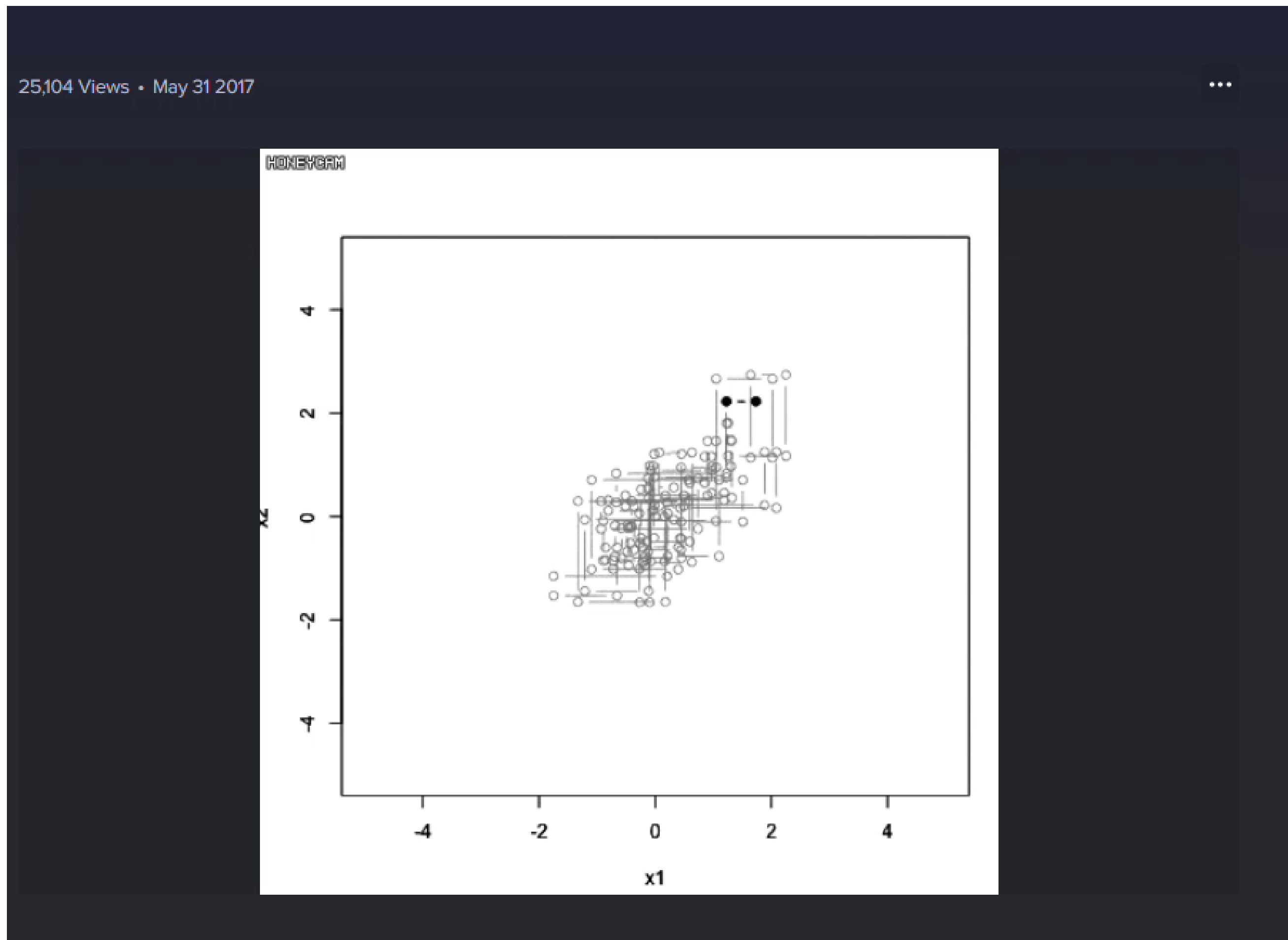
$$\Rightarrow \text{Inv-Gamma} \left(\left(\frac{n}{2} + \alpha \right), \frac{1}{2} \sum_{i=1}^n (x_i - \mu)^2 + \beta \right)$$

Gibbs Sampling - MH 중 하나

- 분포 $P(z)$ 의 불변성 / 각 단계별 + 전체 마르코프 연쇄에 대해서
- 에르고딕성
- 초기 상태의 분포가 명시 돼 있어야함

$$A(\mathbf{z}^*, \mathbf{z}) = \frac{p(\mathbf{z}^*)q_k(\mathbf{z}|\mathbf{z}^*)}{p(\mathbf{z})q_k(\mathbf{z}^*|\mathbf{z})} = \frac{p(z_k^*|\mathbf{z}_{\setminus k}^*)p(\mathbf{z}_{\setminus k}^*)p(z_k|\mathbf{z}_{\setminus k}^*)}{p(z_k|\mathbf{z}_{\setminus k})p(\mathbf{z}_{\setminus k})p(z_k^*|\mathbf{z}_{\setminus k})} = 1$$

Gibbs Sampling



Gibbs Sampling의 취약점 및 보완책

- Random Walking
- Over-relaxation
- Ordered over-relaxation

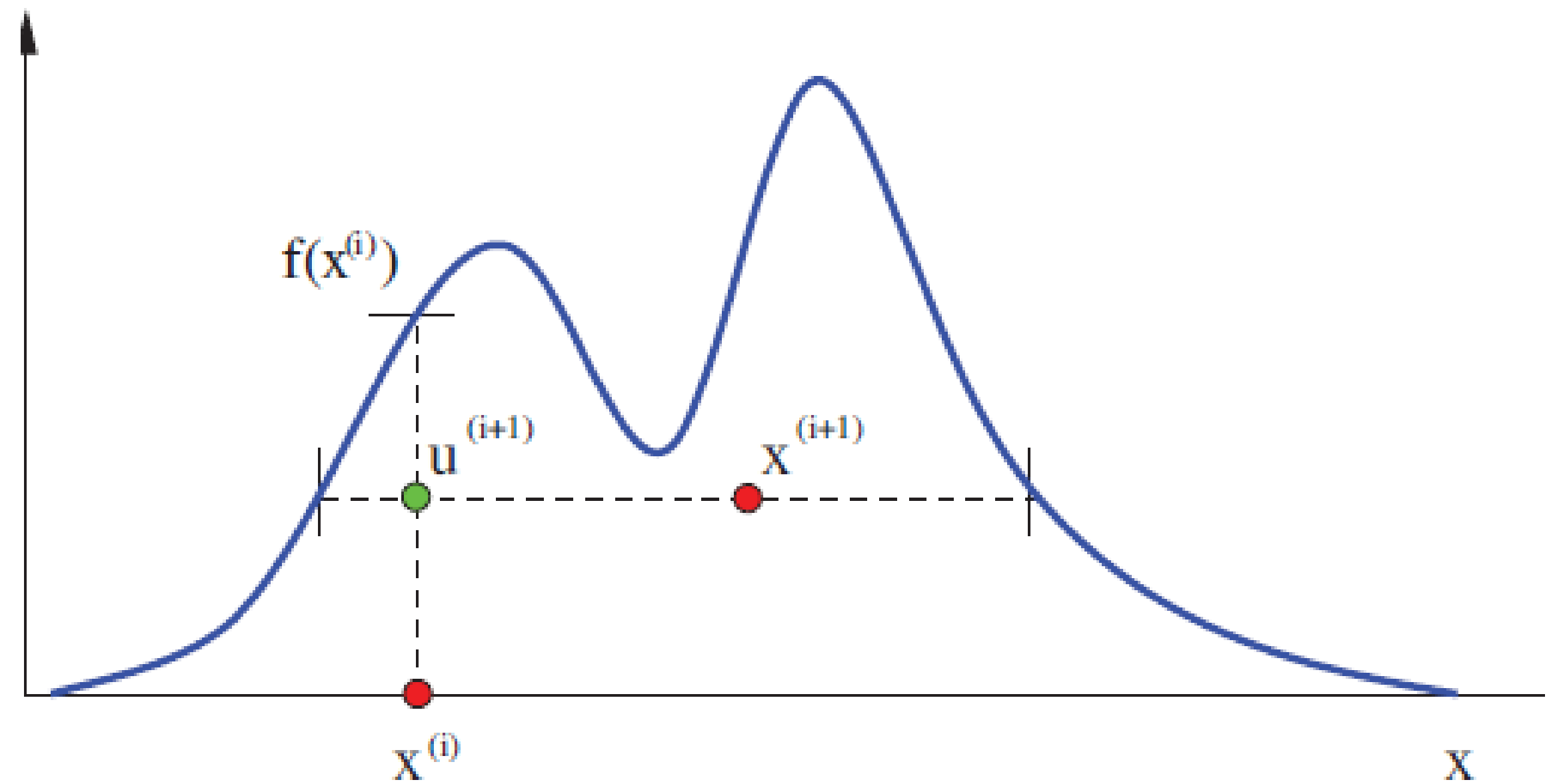
11.4

Slice Sampling

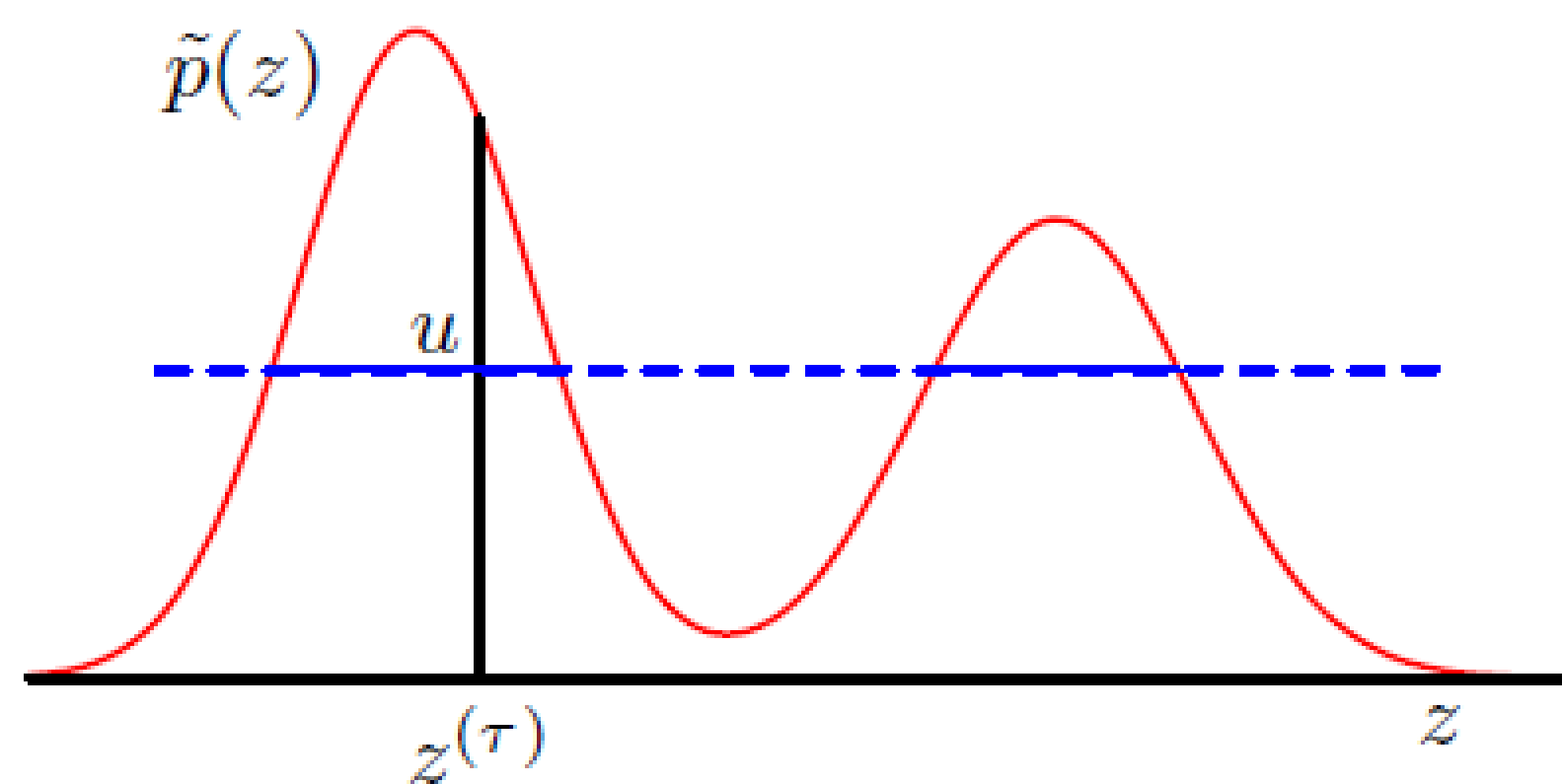
Automatically adjusting step size

Slice Sampling

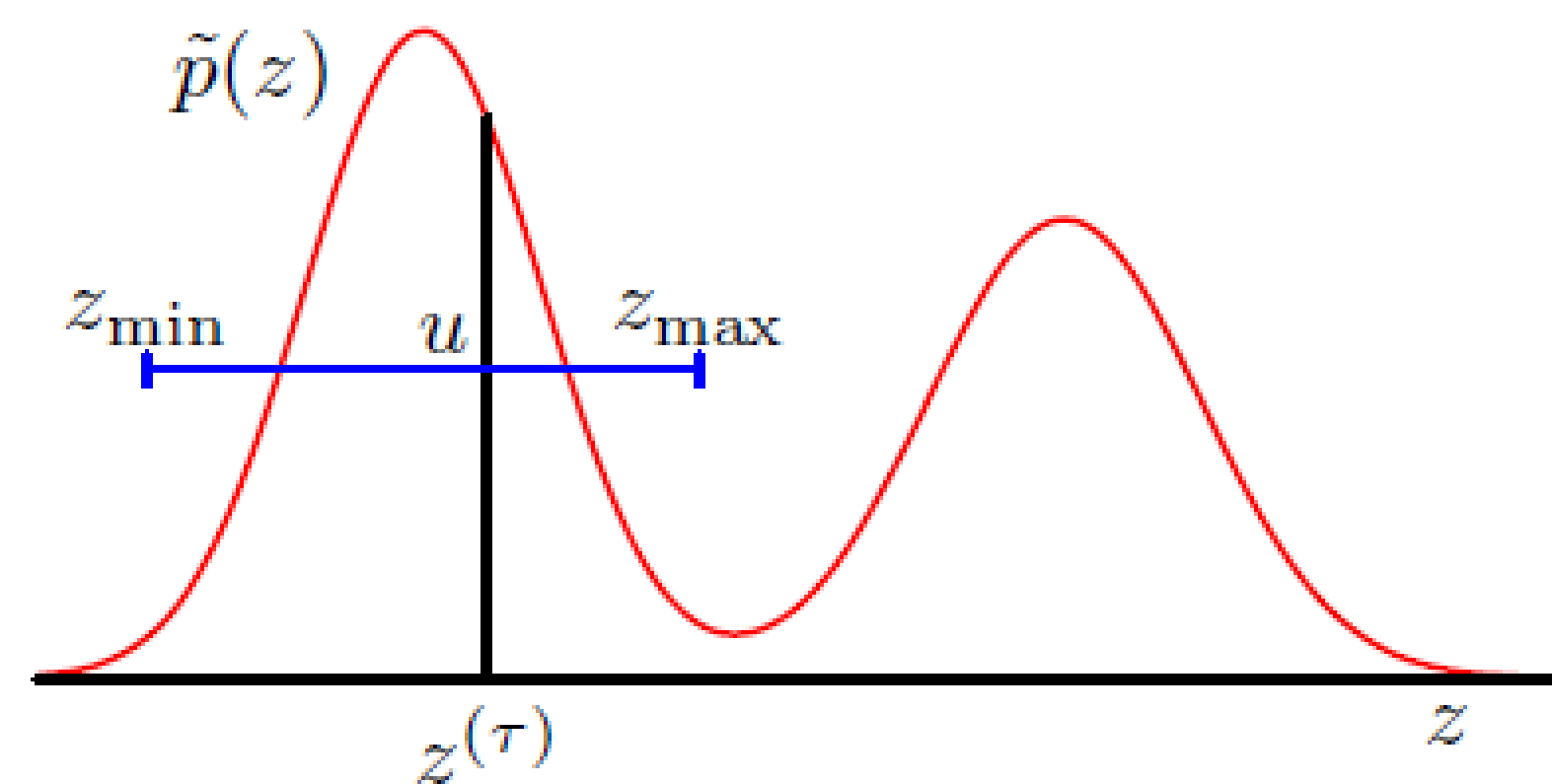
$$p(x, u) = \frac{\mathbf{1}_{\{0 \leq u \leq \tilde{p}(x)\}}}{Z}, \quad \int_0^{\tilde{p}(x)} p(x, u) du = p(x)$$



Slice Sampling



(a)



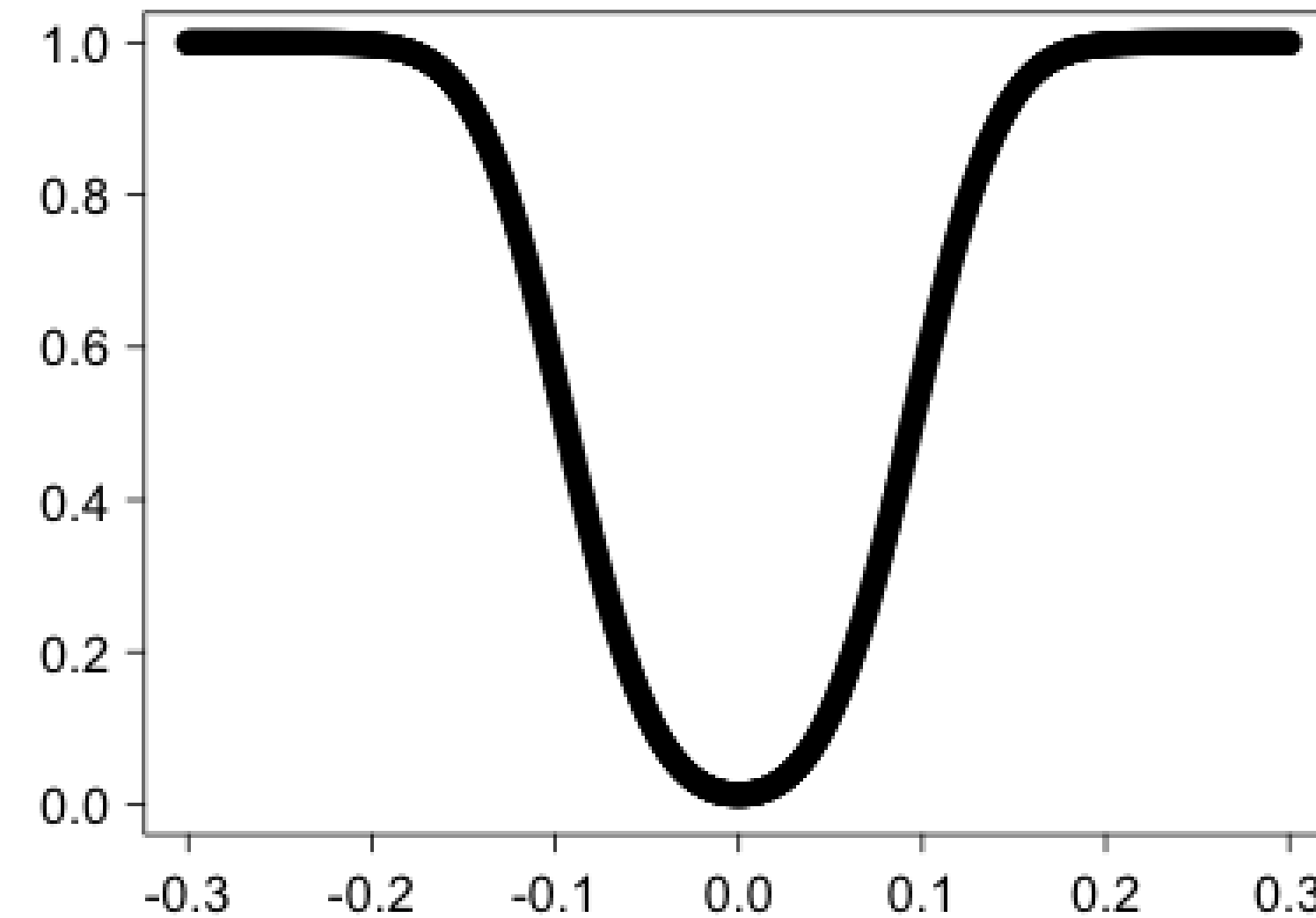
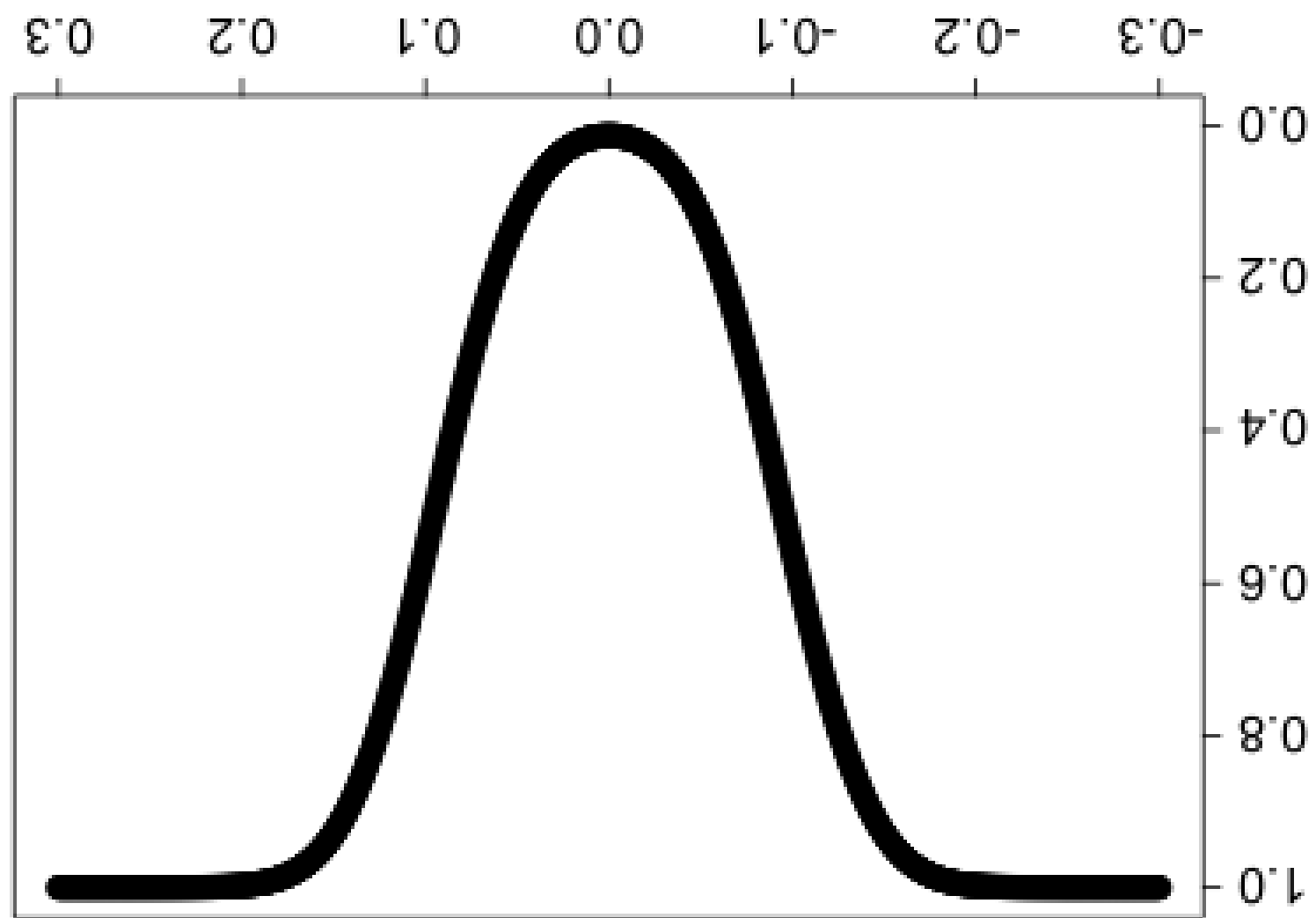
(b)

11.5

Hamiltonian Monte Carlo

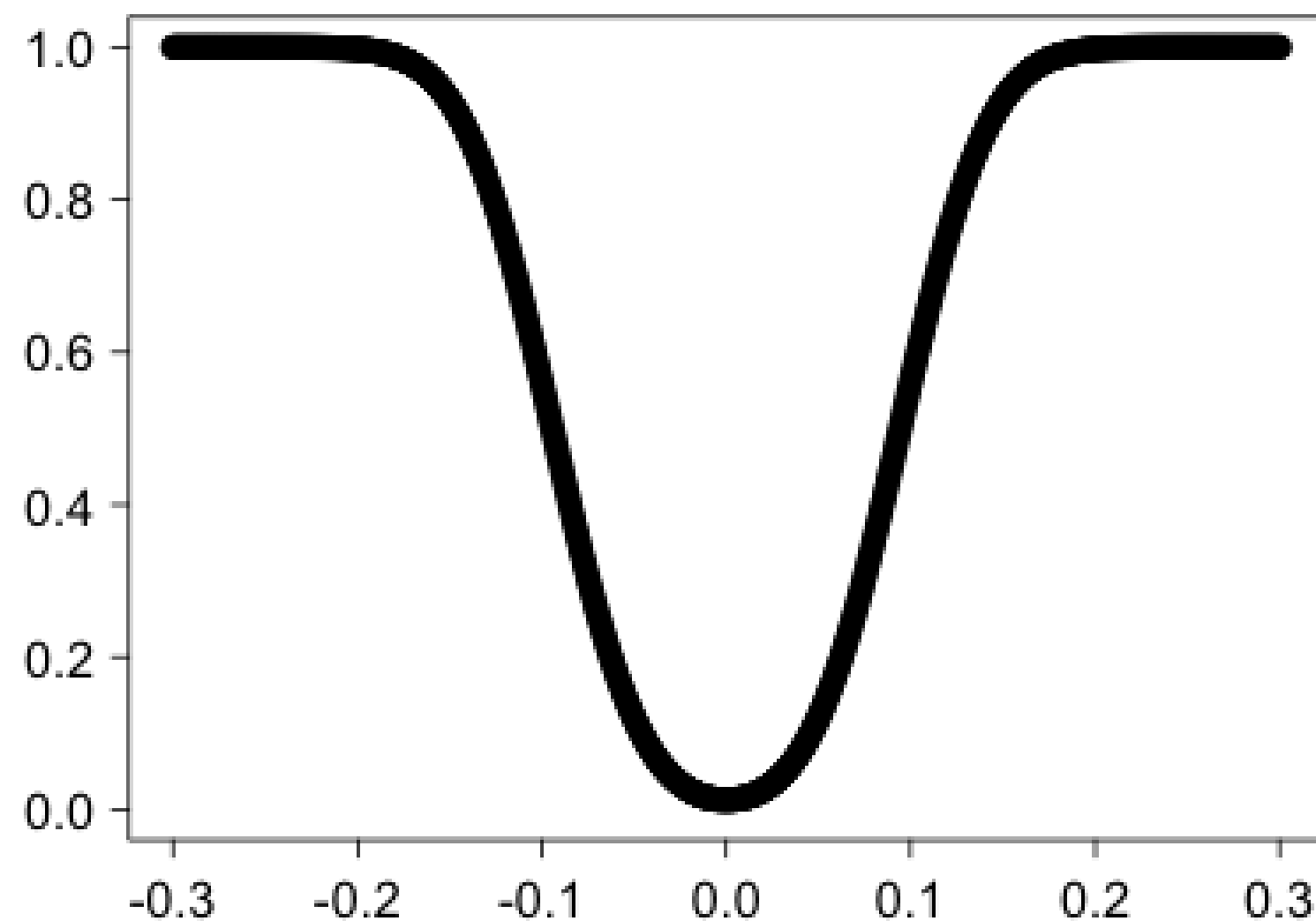
Special case of the Metropolis-Hastings algorithm

Hamiltonian Monte Carlo - Dynamical systems



- Posterior Dist에서 높은 확률이 많이 모여 있는 부분 = 뒤집은 놈에서 가장 아래에 있는 부분

Hamiltonian Monte Carlo - Dynamical systems



롤러코스터를 타는 물리 문제를 만났던 경험을 떠올리자
역학적에너지 보존을 이용해서, 운동량, 각속도를 구했다.

⇔ Y축을 E라고 생각했을 때 E(에너지)가 높아질수록 확률이 낮아짐

$$P(E_i) \propto e^{\frac{-E_i}{T}} \text{ (에너지가 증가할수록 해당 에너지값을 가질 확률이 낮아지는 형태)}$$

Hamiltonian Monte Carlo - Dynamical systems

$$H(x, p) = U(x) + K(p)$$

$$U(x) = -\ln f(x)$$

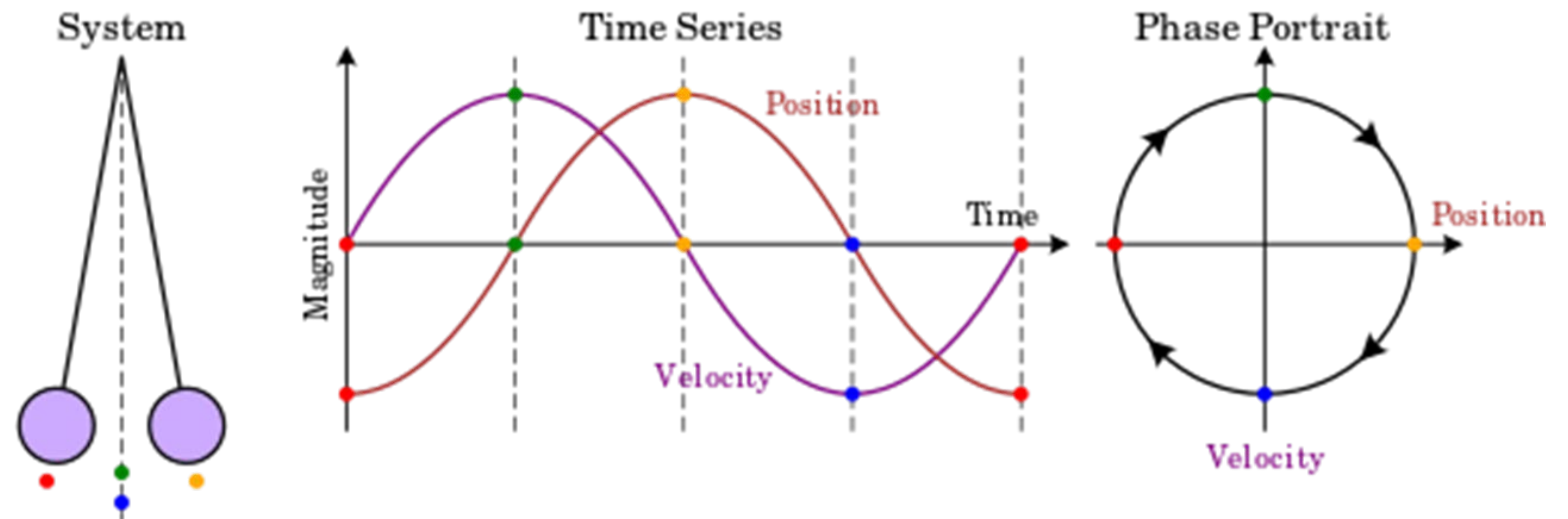
$$V(p) = \sum_{i=1}^d \frac{m^2}{2 \cdot \text{mass}}$$

x : 물체의 "위치(position)"

p : 물체의 "운동량(momentum)"

$U(x)$: 위치 에너지

$K(p)$: 운동 에너지

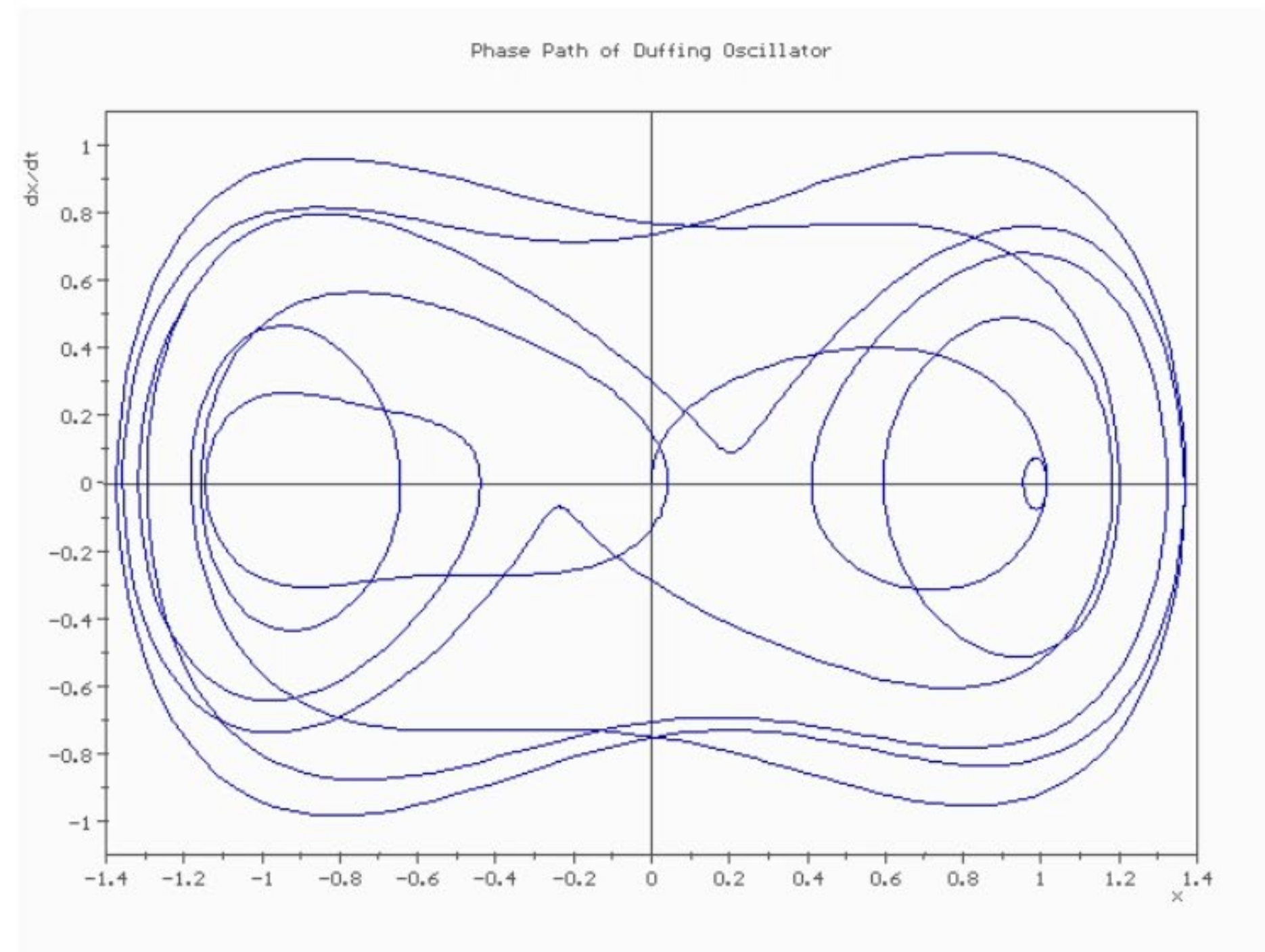


Hamiltonian Monte Carlo - Dynamical systems

$$\begin{aligned}P(x, p) &\propto e^{\frac{-H(x, p)}{T}} \\&= e^{\frac{-U(x) - K(p)}{T}} \\&= e^{\frac{\ln f(x) - \frac{1}{2} p^T M^{-1} p}{T}} \\&\propto f(x) \times -\frac{1}{2} p^T M^{-1} p \\&\equiv f(x) \times -N(p \mid 0, 1)\end{aligned}$$

$$P(x) = \int P(x, p) dp = \frac{1}{Z} f(x) \int N(p \mid 0, 1) dp = \frac{1}{Z} f(x)$$

Hamiltonian Monte Carlo - Dynamical systems



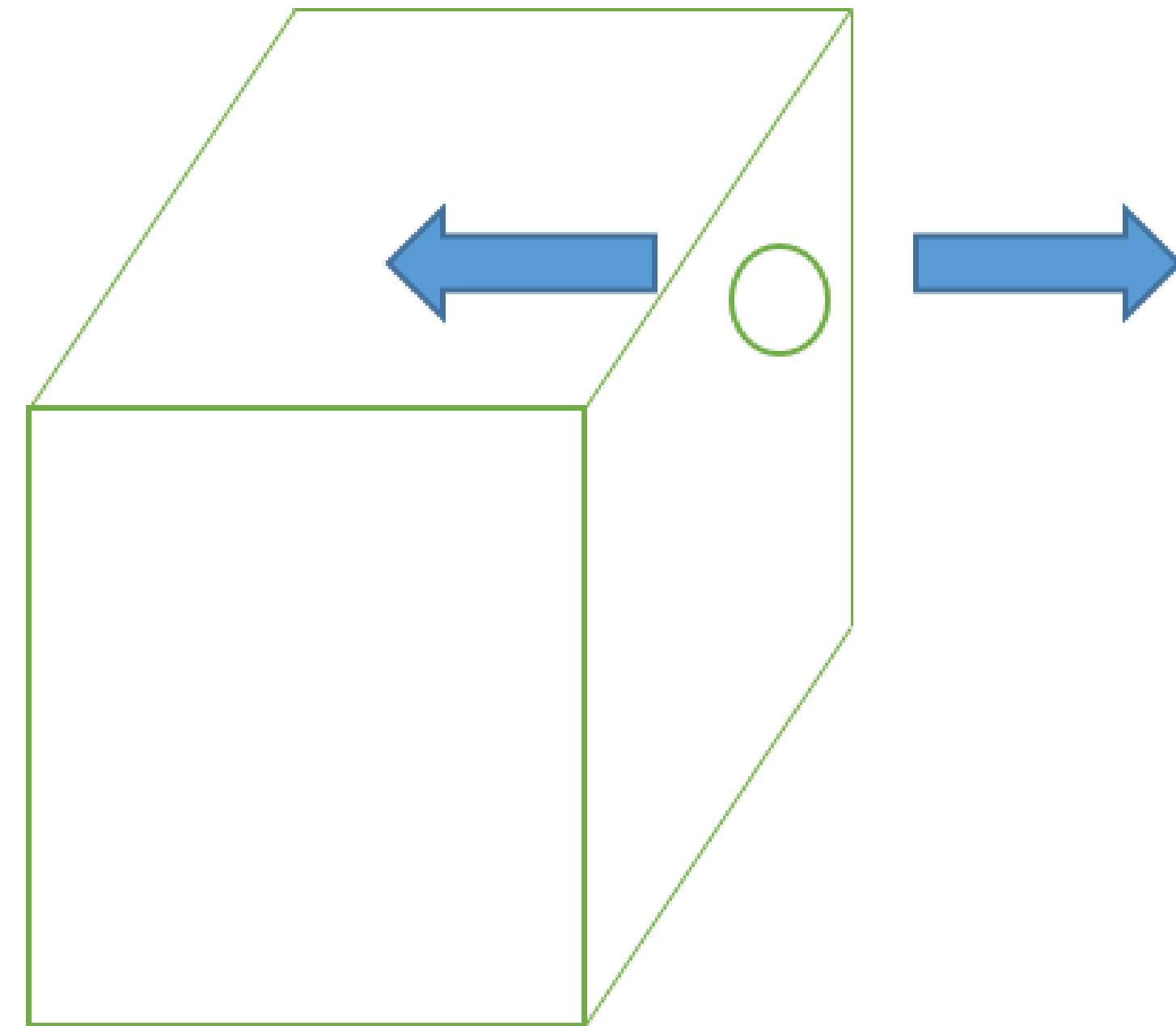
[Pasimi](#) / [CC BY-SA](#) via Wikimedia

Hamiltonian Monte Carlo - Liouville's theorem

1. Divergence Theorem

: 닫힌 평면의 면적분은 밖으로 빠져나오는 Vector Flux의 합을 나타낸다.

$$\text{div } \mathbf{F} = \lim_{V \rightarrow 0} \iint_{S(V)} \frac{\mathbf{F} \cdot \mathbf{n}}{V} dS$$



Hamiltonian Monte Carlo - Liouville's theorem

1. Divergence Theorem

: 닫힌 평면의 면적분은 밖으로 빠져나오는 Vector Flux의 합을 나타낸다.

$$\frac{\partial}{\partial t} \int_{\omega} \rho \, d\omega = - \int_{\omega} \text{div}(\rho \mathbf{v}) \, d\omega$$

$$\int_{\omega} \frac{\partial \rho}{\partial t} + \text{div}(\rho \mathbf{v}) \, d\omega = 0$$

$$\begin{aligned} \text{div}(\rho \mathbf{v}) &= \sum_{i=1}^{3N} \frac{\partial}{\partial x_i} (\rho \dot{x}_i) + \frac{\partial}{\partial p_i} (\rho \dot{p}_i) \\ &= \sum_{i=1}^{3N} \frac{\partial \rho}{\partial x_i} \dot{x}_i + \rho \frac{\partial \dot{x}_i}{\partial x_i} + \frac{\partial \rho}{\partial p_i} \dot{p}_i + \rho \frac{\partial \dot{p}_i}{\partial p_i} \end{aligned}$$

$$\begin{aligned} \text{div}(\rho \mathbf{v}) &= \sum_{i=1}^{3N} \frac{\partial \rho}{\partial x_i} \frac{\partial H}{\partial p_i} + \rho \frac{\partial^2 H}{\partial x_i \partial p_i} - \frac{\partial \rho}{\partial p_i} \frac{\partial H}{\partial x_i} - \rho \frac{\partial^2 H}{\partial p_i \partial x_i} \\ &= \sum_{i=1}^{3N} \frac{\partial \rho}{\partial x_i} \frac{\partial H}{\partial p_i} - \frac{\partial \rho}{\partial p_i} \frac{\partial H}{\partial x_i} \\ &= \{\rho, H\} \end{aligned}$$

$$\frac{\partial \rho}{\partial t} + \{\rho, H\} = 0$$

Hamiltonian Monte Carlo - Liouville's theorem

1. p 를 샘플링한다. $p \sim N(0, 1)$

2. Leap frog steps

- L : leap frog step의 횟수
- Δt : step size

아래의 과정을 총 L 번 반복한다

(2-1) p 를 절반 update한다 : $p \leftarrow p +$

(여기서 ϵ 는 자유롭게 튜닝 가능하다)

(2-2) x 를 update한다 : $x \leftarrow x + \epsilon M$

(2-3) p 의 나머지 절반을 update한다 : $p \leftarrow p + \frac{1}{2}\epsilon \frac{d \ln f(x)}{dx}$

3. Acceptance ratio를 구한다

$$\alpha = \min\left(\frac{f(x^*)P(p^*)}{f(x^{t-1})P(p^{t-1})}, 1\right)$$

4. Acceptance ratio에 따라 이동할지 말지를 결정한다.

$$\theta^t = \begin{cases} x^* & \text{with probability } \alpha \\ x^{t-1} & \text{otherwise} \end{cases}$$

Hamiltonian Monte Carlo - Liouville's theorem

3. Acceptance ratio를 구한다

$$\alpha = \min\left(\frac{f(x^*)P(p^*)}{f(x^{t-1})P(p^{t-1})}, 1\right)$$

4. Acceptance ratio에 따라 이동할지 말지를 결정한다.

$$\theta^t = \begin{cases} x^* & \text{with probability } \alpha \\ x^{t-1} & \text{otherwise} \end{cases}$$

11.6

Estimating the Partition Function

Special case of the Metropolis-Hastings algorithm

Hamiltonian Monte Carlo - Liouville' s thorem

$$p_E(\mathbf{z}) = \frac{1}{Z_E} \exp(-E(\mathbf{z}))$$

$$\begin{aligned} \frac{Z_E}{Z_G} &= \frac{\sum_{\mathbf{z}} \exp(-E(\mathbf{z}))}{\sum_{\mathbf{z}} \exp(-G(\mathbf{z}))} \\ &= \frac{\sum_{\mathbf{z}} \exp(-E(\mathbf{z}) + G(\mathbf{z})) \exp(-G(\mathbf{z}))}{\sum_{\mathbf{z}} \exp(-G(\mathbf{z}))} \\ &= \mathbb{E}_{G(\mathbf{z})} [\exp(-E + G)] \\ &\simeq \frac{1}{L} \sum_l \exp(-E(\mathbf{z}^{(l)}) + G(\mathbf{z}^{(l)})) 1/L \end{aligned}$$

$$\frac{1}{Z_G} \exp(-G(\mathbf{z})) = \frac{1}{L} \sum_{l=1}^L T(\mathbf{z}^{(l)}, \mathbf{z})$$

감사합니다