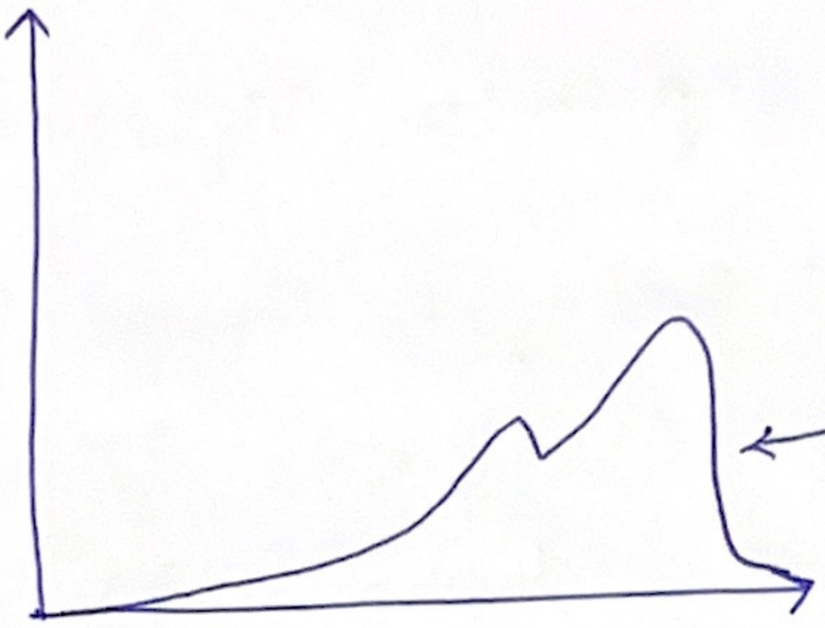
MCMC

- Rejection sampling + Limitations
- Markov Chain
- Examples
- Parameters
- Sample
- Stationary Distribution

Rejection Sampling



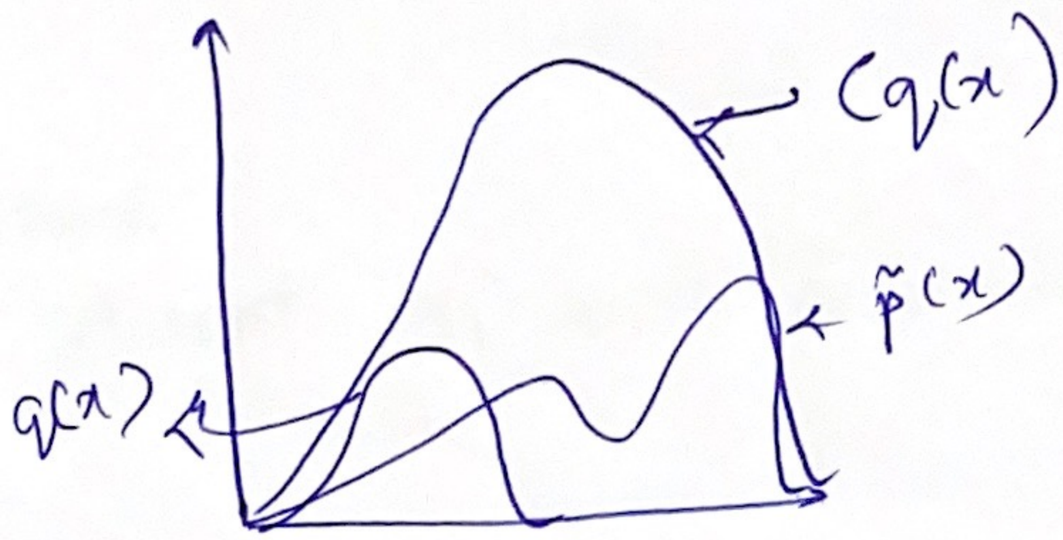
Unnormalized
(e.g. $P(D|\theta) \cdot P(\theta)$)

$$p(x) = \frac{\tilde{p}(x)}{Z_p}$$ ← Normalization constant (e.g. $\int P(D|\theta) P(\theta) d\theta$)

$$Z_p = \int \tilde{p}(x)\, dx$$

1) Proposal : $q(x)$
2) constant : C s.t. $Cq(x) \geq \tilde{p}(x)$

$Cq(x)$ upper envelope of $\tilde{p}(x)$



Steps

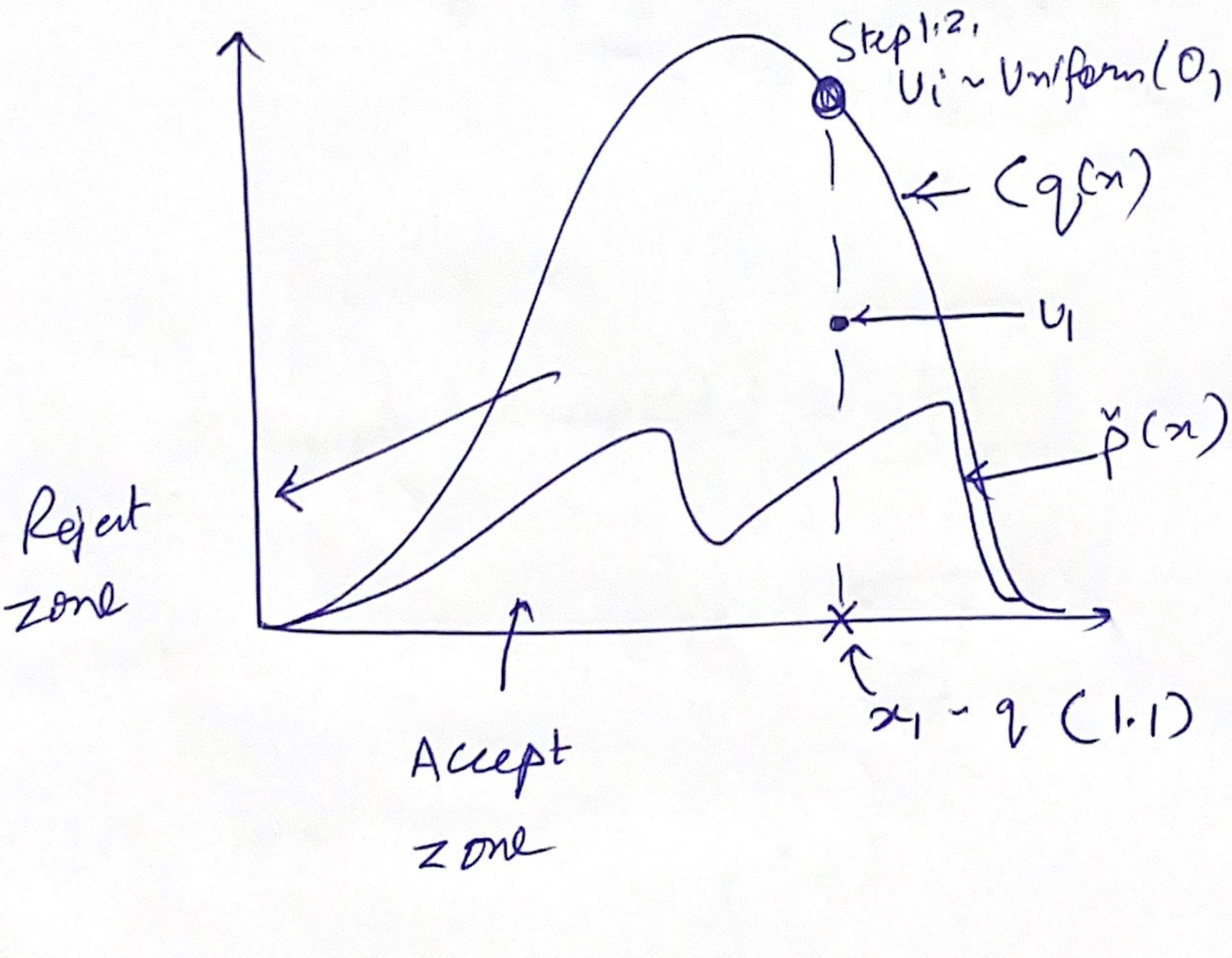
1) For i in [1, ... N_tries]:

1.1). Sample $x_i \sim q$

1.2) Sample $u_i \sim \text{UNIF}(0, Cq(x_i))$

1.3) If $u_i > \tilde{p}(x_i)$: REJET

1.4) Else : accept



OR

1.1) same

1.2) Accept x_i with probability $\frac{\tilde{p}(x_i)}{Cq(x_i)}$

$$P(\text{Sample} \mid \text{Accepted}) = \frac{P(\text{Accept} \mid \text{Sample})\, P(\text{Sample})}{P(\text{Accept})}$$

$$= \frac{\left\{\frac{\tilde{p}(x)}{c\,q(x)}\right\} \times q(x)}{P(\text{Accept})}$$

$$P(\text{Accept}) = \int \frac{\tilde{p}(x)\, q(x)}{c\, q(x)}\, dx = \frac{1}{c}\int \tilde{p}(x)\, dx$$

$$\therefore P(\text{Sample} \mid \text{Accepted}) = \frac{\tilde{p}(x)}{\frac{1}{c}} = p(x)$$

Limitations

1) $P(\text{Accept}) \propto \frac{1}{C}$

Wasteful for large C

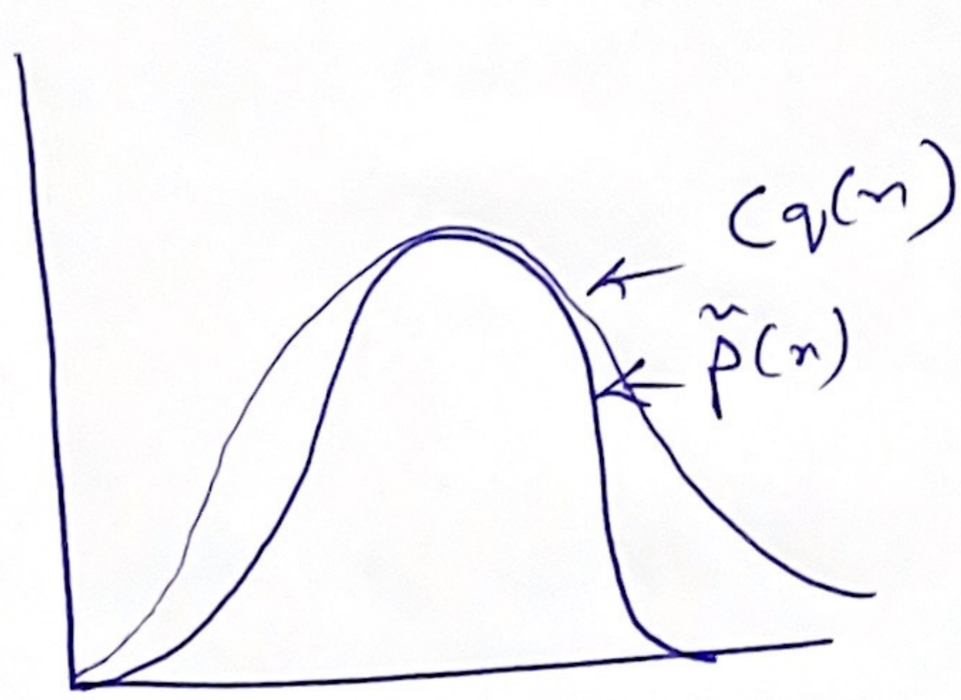
2) In high dimensions

Assume $\tilde{p}(x) = N(0, \sigma_p^2 I)$

$q(x) = N(0, \sigma_q^2 I)$

$\sigma_q > \& \sigma_p$

$C_{\text{optimum}} = ? = \frac{\tilde{p}(0)}{q(0)} = \left(\frac{\sigma_q}{\sigma_p}\right)^D$



At $\sigma_q/\sigma_p = 1.1$; $D = 100$; $P(\text{Accept}) = \frac{1}{C}$

$= 10^{-4}$

3) I. I. D. samples?!



Would it make sense to also have samples

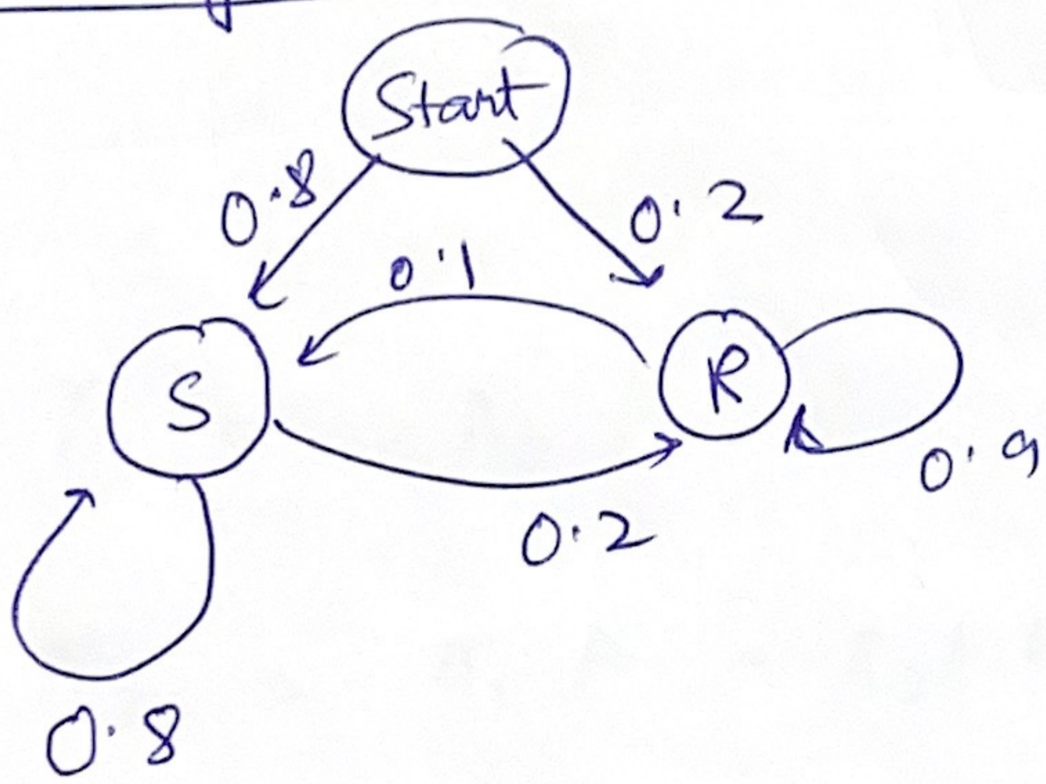
$x_1 + \Delta x$; $x_1 - \Delta x$

or $N(x_1, \sigma)$?

Markov Chain

http://nipunbatra.github.io/hmm

Stationary distribution



$\pi_t(j) = P(S_j \text{ at time } t)$

$\pi_0(S) = 0.8$; $\pi_0(R) = 0.2$

$\pi_1(S) = 0.8 \times 0.8 + 0.2 \times 0.1 = .66$

$\pi_1(R) = 0.2 \times .9 + 0.8 \times 0.2 = .34$

$\pi_2(S) = \pi_1(S) * 0.8 + \pi_1(R) \times 0.1 = .562$

$\pi_2(R) = 0.438$

$$\pi_0 = \begin{bmatrix} .8 & .2 \end{bmatrix}$$

$$A = \begin{matrix} \\ S \\ R \end{matrix} \overset{\overset{t+1}{\downarrow}}{\begin{matrix} S & R \end{matrix}} \begin{bmatrix} .8 & .2 \\ .1 & .9 \end{bmatrix}$$

$$\pi_1 = \pi_0 A$$

$$\pi_2 = \pi_1 A = \pi_0 A \cdot A$$

or

$$\pi_N = \pi_{N-1} \pi_0 A^N$$

Stationarity

$$\pi_N = \pi_{N+1}$$

or

$$\pi_N = \pi_N A$$

$$\underset{1\times 2}{[\pi_S \quad \pi_R]}\underset{2\times 2}{\begin{bmatrix} .8 & .2 \\ .1 & .9 \end{bmatrix}} = \begin{bmatrix} .8\pi_S + .1\pi_R, \\ .2\pi_S + .9\pi_R \end{bmatrix}$$

$\therefore \quad .8\pi_S + .1\pi_R = \pi_S$

& $\quad .2\pi_S + .9\pi_R = \pi_R$

$\Rightarrow \quad .2\pi_S = .1\pi_R \quad$ or $\pi_R = 2\pi_S$

$\pi_R + \pi_S = 1$

or $3\pi_S = 1 \quad$ or $\pi_S = \frac{1}{3}$

$\pi_R = \frac{2}{3}$

$\pi = [2/3 \quad 1/3] \quad \pi = [1/3 \quad 2/3]$

Verify $[1/3 \quad 2/3]\begin{bmatrix} .8 & .2 \\ .1 & .9 \end{bmatrix}$

$= \left[\frac{.8}{3} + \frac{.2}{3}, \quad \frac{.2}{3} + \frac{.9.8}{3}\right] = [1/3 \quad 2/3]$

Metropolis Algorithm

Goal: Find samples of $\tilde{p}(x)$

1. Propose initial sample x_0

2. For i in [0, ... N_samples]:

2.1. Draw $x' \sim q(x'|x_t)$

e.g. $q \sim N(x'; x_t; \sigma^2 = 1)$
(Markov chain)

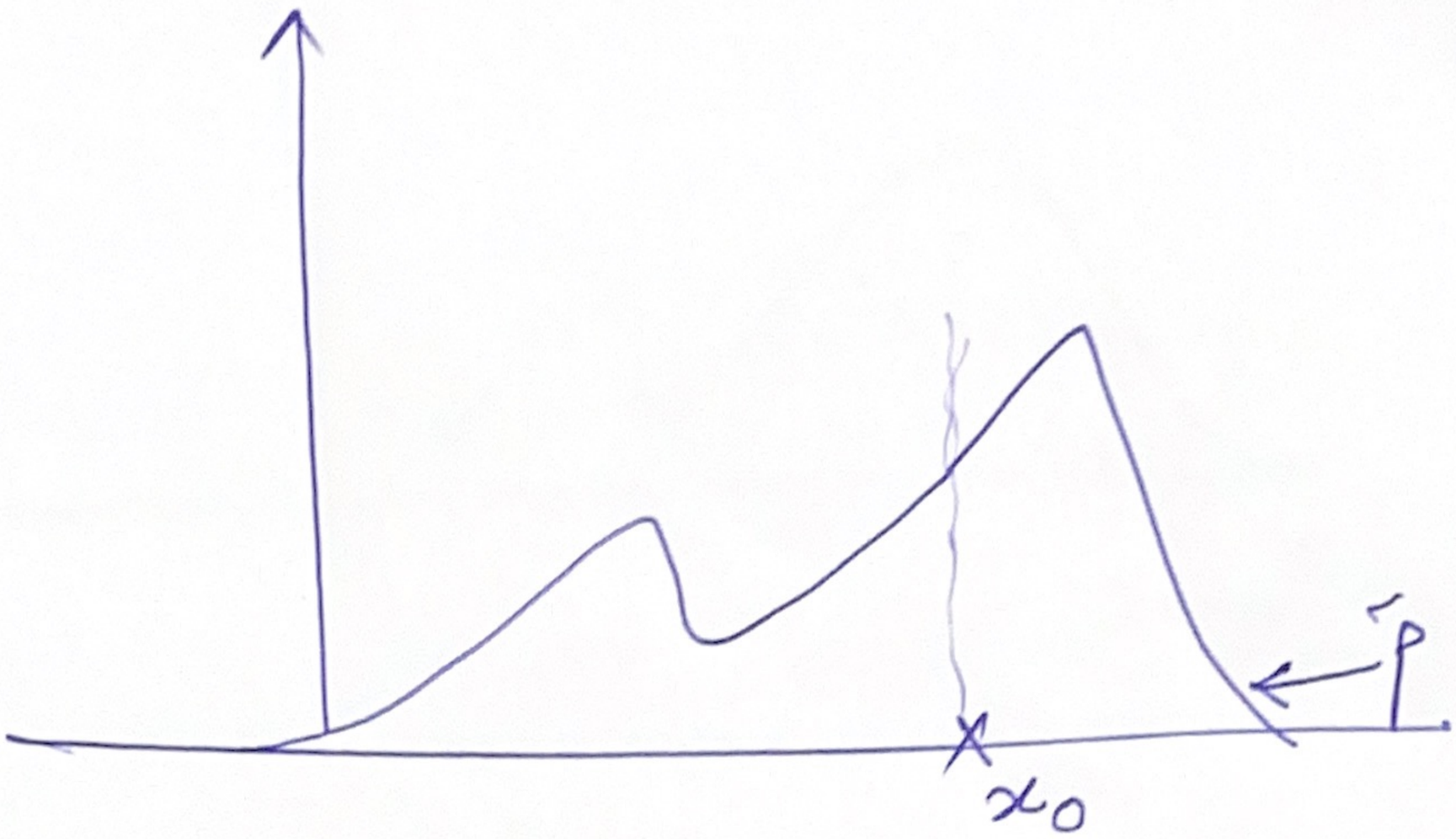
2.2. $a = \frac{\tilde{p}(x')}{\tilde{p}(x_t)}$

2.3. if $a \geq 1$; $x_{t+1} = x'$

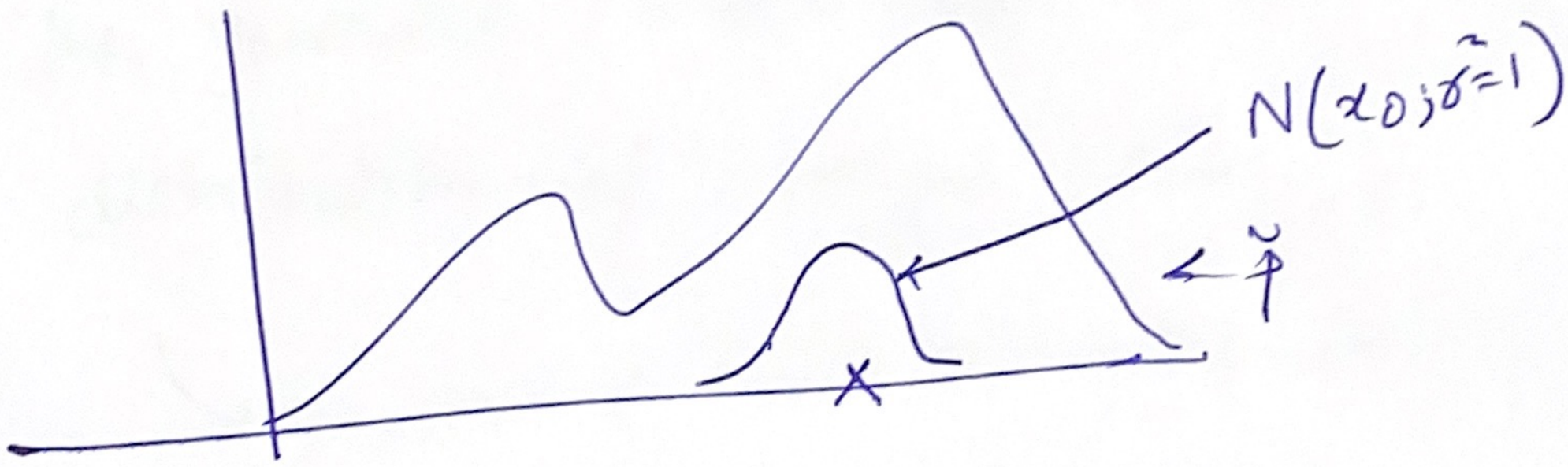
2.4. else:

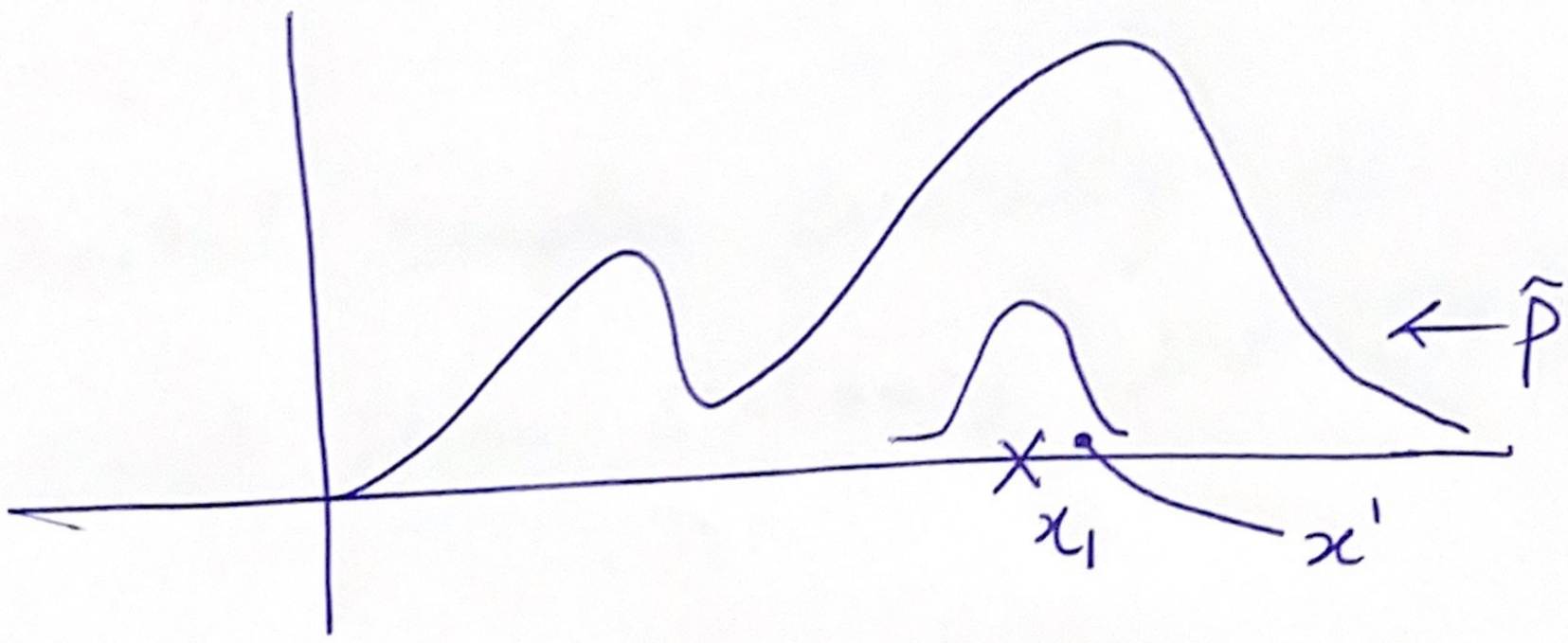
$x_{t+1} \leftarrow x'$ w/ probability 'a'

$x_{t+1} \leftarrow x_t$ w/ probability 1−a

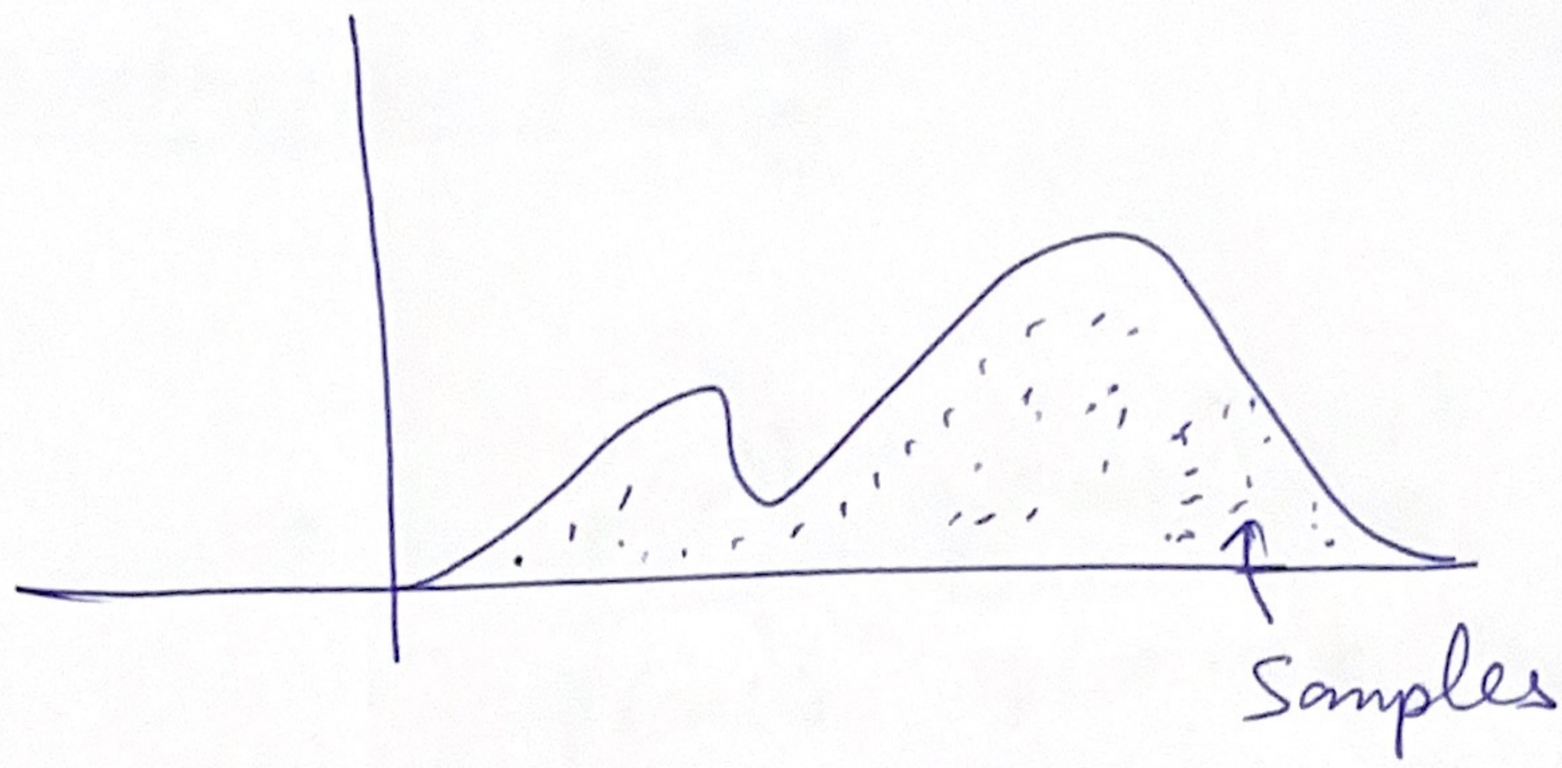


$$q \sim N(x'; x_t; \sigma^2 = 1)$$





$$a = \frac{\hat{p}(x')}{\tilde{p}(x_1)} \geq 1 \quad \therefore \quad x_2 = x'$$



1) Density of x_t approaches $p(x)$ as $t \to \infty$

2) The generated Markov chain has stationary distribution of target density.

↳ Fraction of time we spend in state $x \propto p(x)$

3) Initial samples generally discarded (not from stationary distribution)

Metropolis Hasting

$$a = \frac{\tilde{p}(x')}{\tilde{p}(x_t)} \frac{q(x_t | x')}{q(x' | x_t)}$$

I.I.D. or Correlated Samples

* Thinning – Consider every k^{th} sample from MC

* Practical tip: Burn-In: Ignore first 'S' Samples as chain has not reached stationarity.