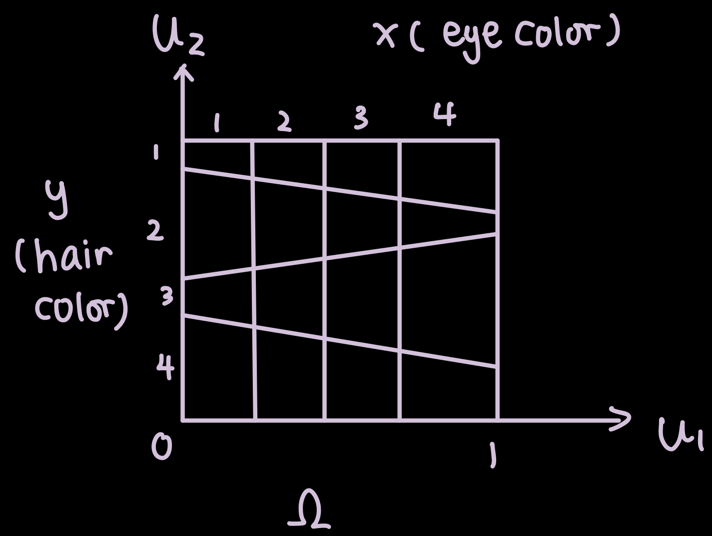
Gibbs Sampler

(Iterative Conditional Sampler)

Background in Probability (conditioning)

Discrete Case



Random Sample (or throw into) a pair in Ω

$P(x,y) =$ area of cell (x,y)

$P_X(x) =$ area of $(x) = \sum_y P(x,y)$

$P_Y(y) = \sum_x P(x,y)$

Operation ① : marginalization

Operation ② : Conditioning

$$P_{Y|X}(y|x) = \frac{P(x,y)}{P_X(x)} \quad \frac{\text{Joint}}{\text{marginal}}$$

randomly throw into (x)

change sample space to (x)

Operation ③: Factorization

$$P(x,y) = P(x)\,P(y|x) = P(y)\,P(x|y)$$

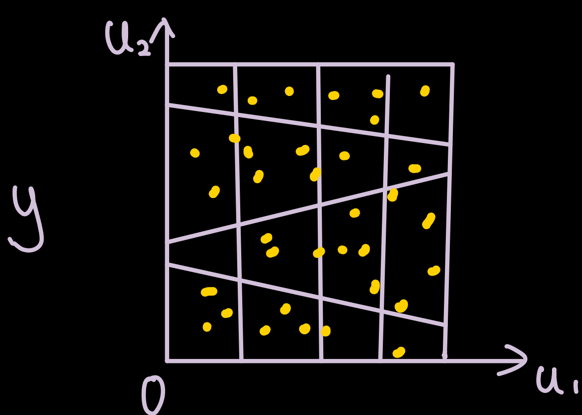
$P(x,y) \downarrow$ area (x, y)

$P(x) \downarrow$ area of (x)

$P(y|x) \downarrow$ $\dfrac{\text{area of } (x\ y)}{\text{area of } (x)}$

A lot of machine learning algorithms are based on these three operations.

New discrete case



Repeat a large # of trees

$P(x,y)$ = how often fall into (x,y)

$P(x)$ = - - - - - (x) → frequency

$P(y|x)$ = when fall into x, how often also fall into y

↳ relative frequency

$P(\text{alarm} \mid \text{fire}) = 1 \qquad P(\text{fire} \mid \text{alarm}) \approx 0$

Back to Gibbs Sampler

Target : $P(x, y)$

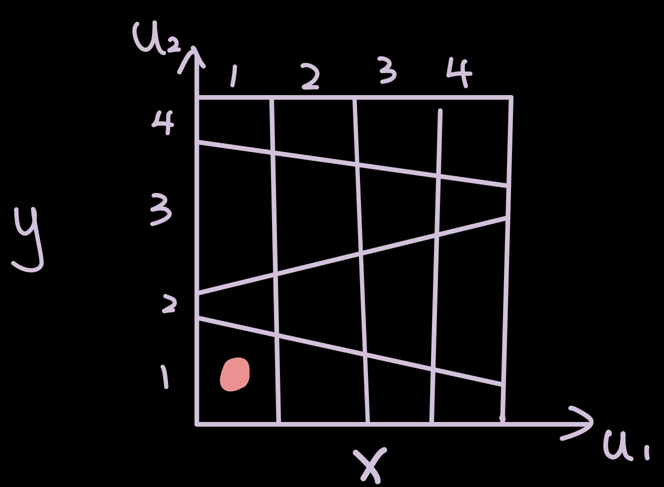
Algorithm : Start from (x_0, y_0)

Iterative, $t = 1, 2 \ldots, T$

Step 1 : Sample $x_t \sim P(x \mid y_{t-1})$

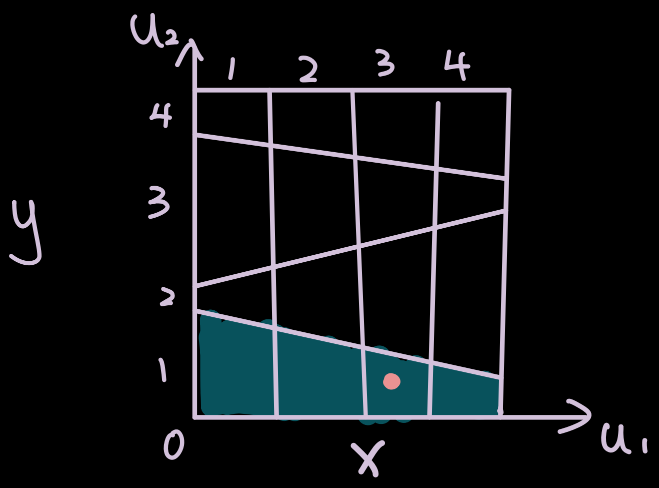
Step 2 : Sample $y_t \sim (y \mid x_t)$

Example:



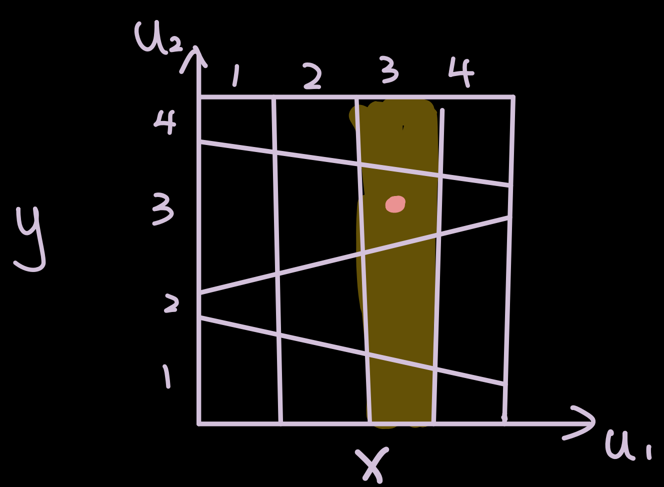
$(x_0 = 1,\ y_0 = 1)$

$x_1 \sim P(x \mid y = 1) \longrightarrow$ randomly relocate in row $(y=1)$ ...



now $x_1 = 3$:

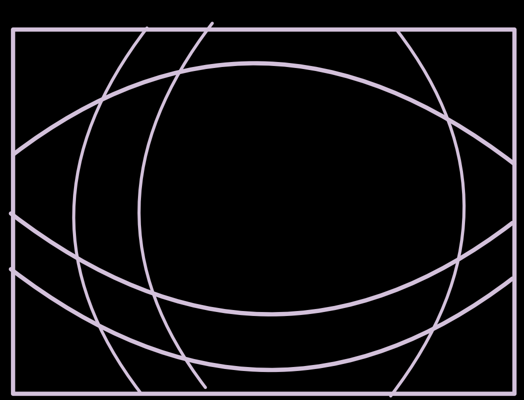
$y_1 \sim P(y \mid x=3)$ → randomly relocate in column $(x=3)$ . . .



now $y_1 = 3$

$\xrightarrow[\infty]{t}$ your location becomes a random point in Ω

Example:

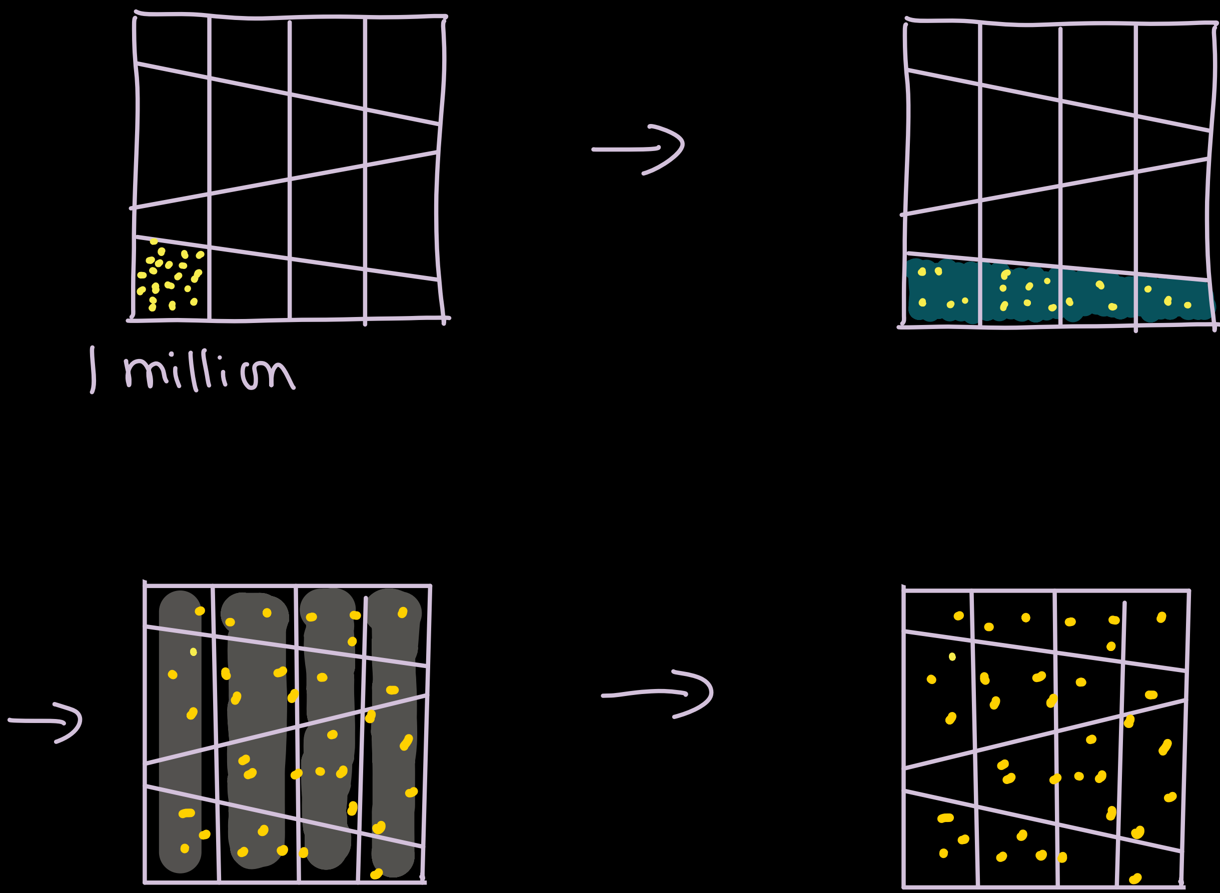


Replace sampling by integration

→ local maximum of $P(x,y)$

"coordinate ascent"

Population migration



Uniform over Ω

$P(x, y)$

$\downarrow$

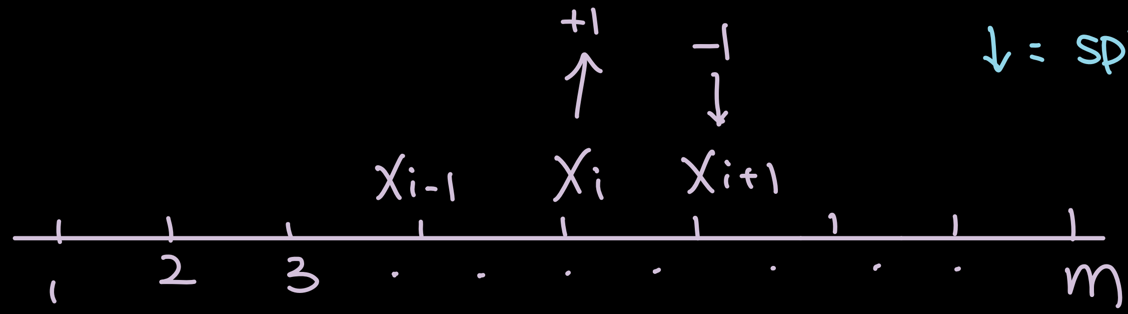
Stationary equilibrium

Ising model

1D space

↑ = spin up

↓ = spin down



$$X = (X_1 \; X_2 \; \ldots \; X_i \; \ldots \; X_m)$$

energy function

$$\mathcal{E}(x) = -(X_1X_2 + X_2X_3 + \cdots + X_{i-1}X_i + X_iX_{i+1} + \cdots + X_{m-1}X_m)$$

gibbs dist

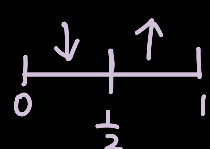
$$\pi(X) = \frac{1}{Z} e^{-\frac{\mathcal{E}(x)}{T}} = \frac{1}{Z} e^{\beta \sum_{i=1}^{m-1} X_i X_{i+1}}$$

"Ferromagnetism" , where $\beta = 1/T$

Pseudo code:

Initialize $X = (X_1 \ldots X_i \ldots X_m)$

$X_i \overset{iid}{\sim}$ Bernoulli $(\frac{1}{2})$ fair coin



Iterate: for (i in 1 : m)

Sample $X_i \sim \pi(X_i \mid X_{[-i]})$

current values of $X_1 \ldots X_{i-1} \; X_{i+1} \ldots X_m$

# of possible $x = 2^m$ local / small changes. feasible

$$\pi(X_i \mid X_{-i}) \xrightarrow{\text{Conditioning}} = \frac{\pi(X_i, X_{-i})}{\pi(X_{-i})}$$

$$\pi(X_{-i}) \xrightarrow{\text{marginalization}} \pi(X_i = +1, X_{-i}) + \pi(X_i = -1, X_{-i})$$

$$\pi(X_i, X_{-i}) = \frac{1}{Z} e^{\beta \sum_{i=1}^{m-1} X_i X_{i+1}}$$

then:

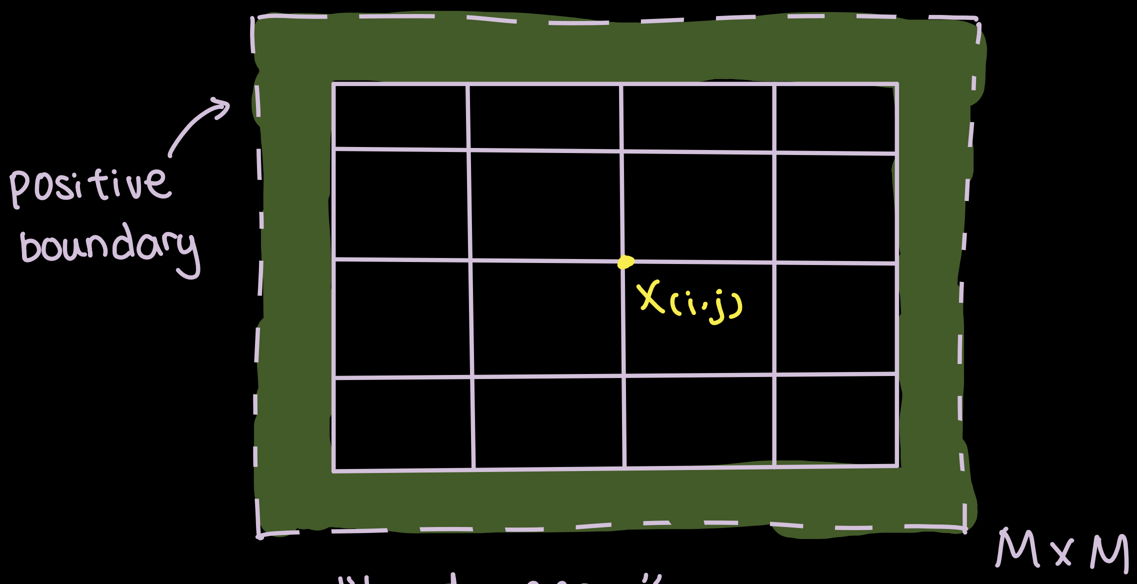
$$= \frac{\cancel{\frac{1}{Z}} e^{\beta \sum_{i=1}^{m-1} X_i X_{i+1}}}{\cancel{\frac{1}{Z}} e^{\beta \underbrace{\sum_{k=1}^{m-1} X_k X_{k+1}}_{X_i = +1}} + \cancel{\frac{1}{Z}} e^{\beta \underbrace{\sum_{k=1}^{m-1} X_k X_{k+1}}_{X_i = -1}}}$$

$$\pi(X_i = +1 \mid X_{-i}) = \frac{e^{\beta(X_{i-1} + X_{i+1})}}{e^{\beta(X_{i-1} + X_{i+1})} + e^{-\beta(X_{i-1} + X_{i+1})}}$$

$$= \pi(X_i = +1 \mid X_{i-1}, X_{i+1}) = \rho$$

only depends on the spin of its two neighbors

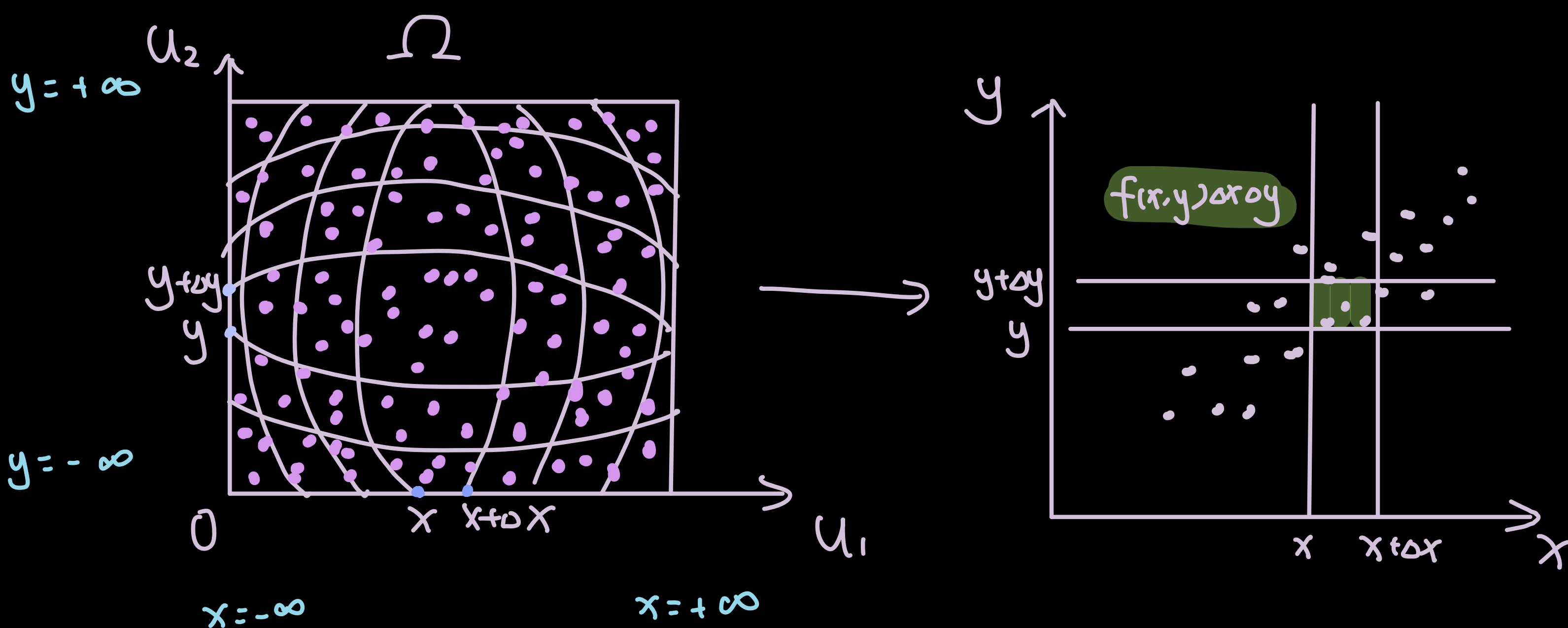
2D space



"heat map"

spontaneous magnitization : Iron

Continuous Case



U-space

xy space

① Marginalization : $f_X(x)\cancel{\Delta x} = \sum_y f(x,y)\cancel{\Delta x}\Delta y$

$$= \int f(x,y)\,dy$$

- $f_Y(y) = \int f(x,y)\,dx$

② Conditioning:

$$f_{Y|X}(y|x)\,\Delta y = \frac{f(x,y)\,\Delta x\,\Delta y}{f_X(x)\,\Delta x}$$

$$P(Y \in (y, y+\Delta y) \mid X \in (x, x+\Delta x))$$

- $f_{X|Y}(x|y) = \dfrac{f(x,y)}{f_Y(y)}$

③ Factorization:

$$f(x,y) = f(x)\,f(y|x)$$
$$= f(y)\,f(x|y)$$

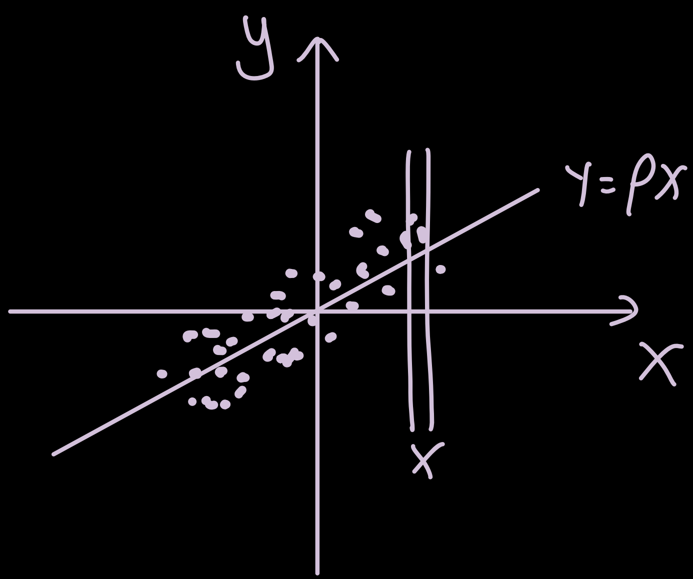
Bivariate Normal

$$X \sim N(0,1) \qquad f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

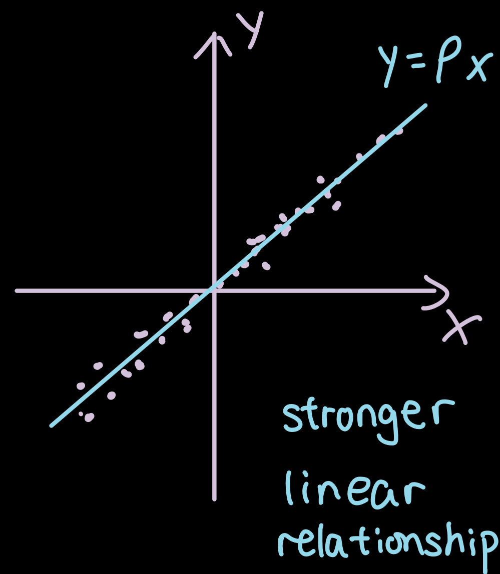
$$[Y \mid X = x] \sim N(\underbrace{\rho x}_{\mu}, \underbrace{1-\rho^2}_{\sigma^2})$$

$$f(y|x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$

$$= \frac{1}{\sqrt{2\pi(1-\rho^2)}} e^{-\frac{(y-\rho x)^2}{2(1-\rho^2)}}$$



as $\rho \to 1$



$$f(x,y) = f(x)\, f(y|x)$$

$$= \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2}\left(x^2 + \frac{(y-\rho x)^2}{1-\rho^2}\right)}$$

$$\frac{x^2 - \cancel{\rho^2 x^2} + y^2 - 2\rho xy + \cancel{\rho^2 x^2}}{1-\rho^2}$$

$$= \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{x^2+y^2-2\rho xy}{2(1-\rho^2)}}$$

$\xrightarrow{\text{symmetry}}$ $f(x|y)\ [X \mid Y=y] \sim N(\rho y,\ 1-\rho^2)$

Using Gibbs Sampler

Initialize X_0 Y_0

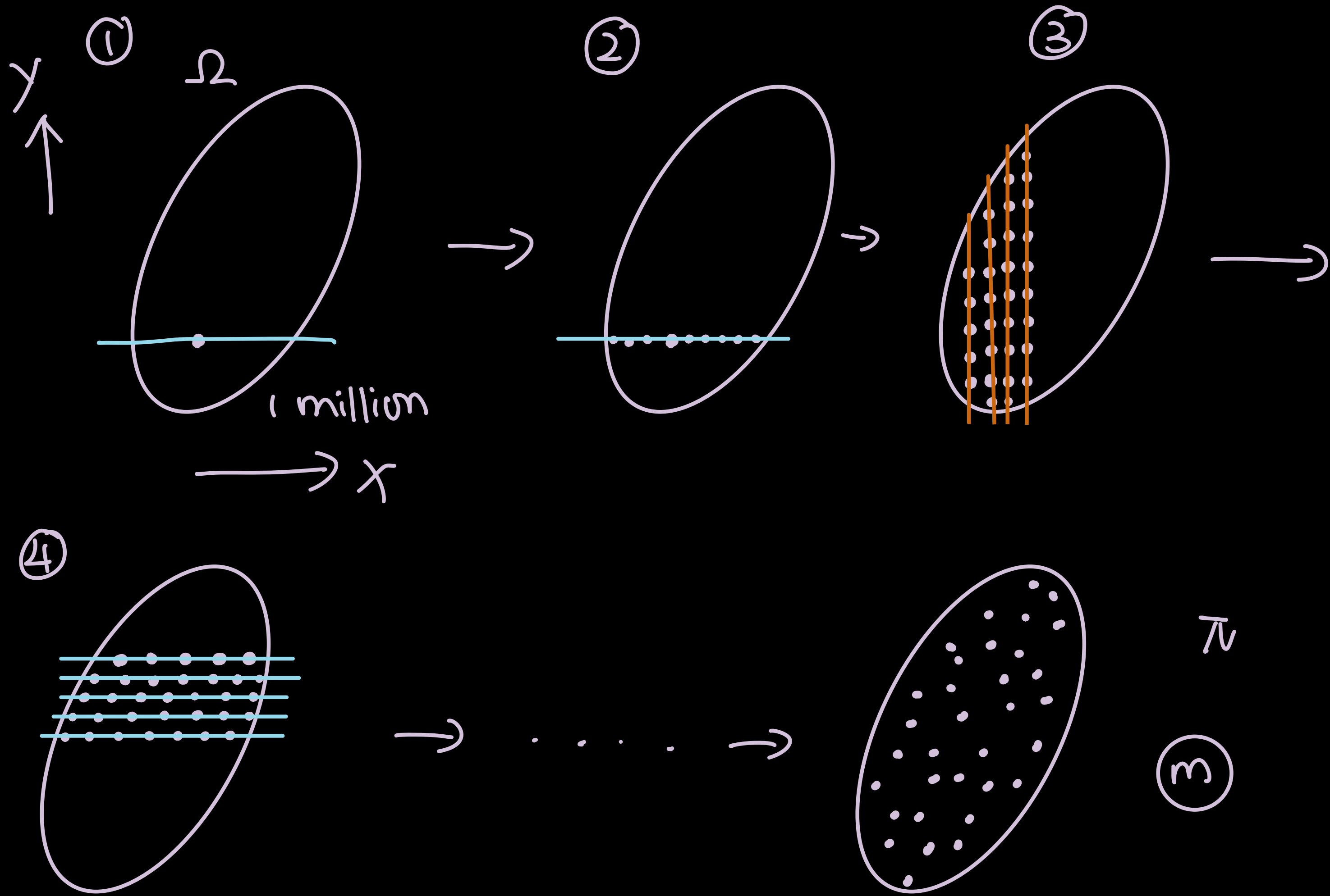
Iterate:

$$X_t \sim N(\rho Y_{t-1}, 1-\rho^2) = \rho Y_{t-1} + \sqrt{1-\rho^2}\, N(0,1)$$

$$Y_t \sim N(\rho X_t, 1-\rho^2) = \rho X_t + \sqrt{1-\rho^2}\, N(0,1)$$

Intuitive Example

$\pi(x,y) \sim \text{Unif}(\Omega)$



Monte-Carlo Integration

$$I = E_{\pi}\left(h(x)\right)$$

$$\text{MCMC} \longrightarrow \underbrace{X_0 \ X_1 \ X_2 \cdots}_{\text{Burn-in period}} X_B \ \overset{\sim \pi}{X_{B+1}} \cdots \overset{\sim \pi}{X_T}$$

$$\hat{I} = \frac{1}{T-B} \sum_{i=B+1}^{T} h(X_i)$$

$$E(\hat{I}) = I \qquad Var(\hat{I})$$

Bayesian Statistics

parameter $\theta \sim P(\theta)$ prior distribution

$[\underset{\downarrow\ \text{data}}{x} | \theta] \sim p(x|\theta)$

$\rightarrow$ posterior $p(\theta|x) \xRightarrow{\text{conditioning}} \dfrac{P(\theta, x)}{P(x)} \; = $

factorization

marginalization

$$= \frac{P(\theta)P(x|\theta)}{\underbrace{\int P(\theta)P(x|\theta)\,d\theta}_{\rightarrow Z}}$$

Use MCMC $\sim p(\theta | x)$

gives us plausible values of θ