Monte-Carlo Markov Chain

When i.i.d. Monte-Carlo is too hard.

Can't sample from f directly on sample space Ω

But there is a conditional distribution

$$q(\underset{\sim}{x}_{t+1} \mid \underset{\sim}{x}_t)$$

So starting at some $\underset{\sim}{x}_0$

and adding $\underset{\sim}{x}_1 \sim q(\cdot \mid x_0)$

$$\vdots$$

$$\underset{\sim}{x}_{t+1} \sim q(\cdot \mid \underset{\sim}{x}_t)$$

$$\vdots$$

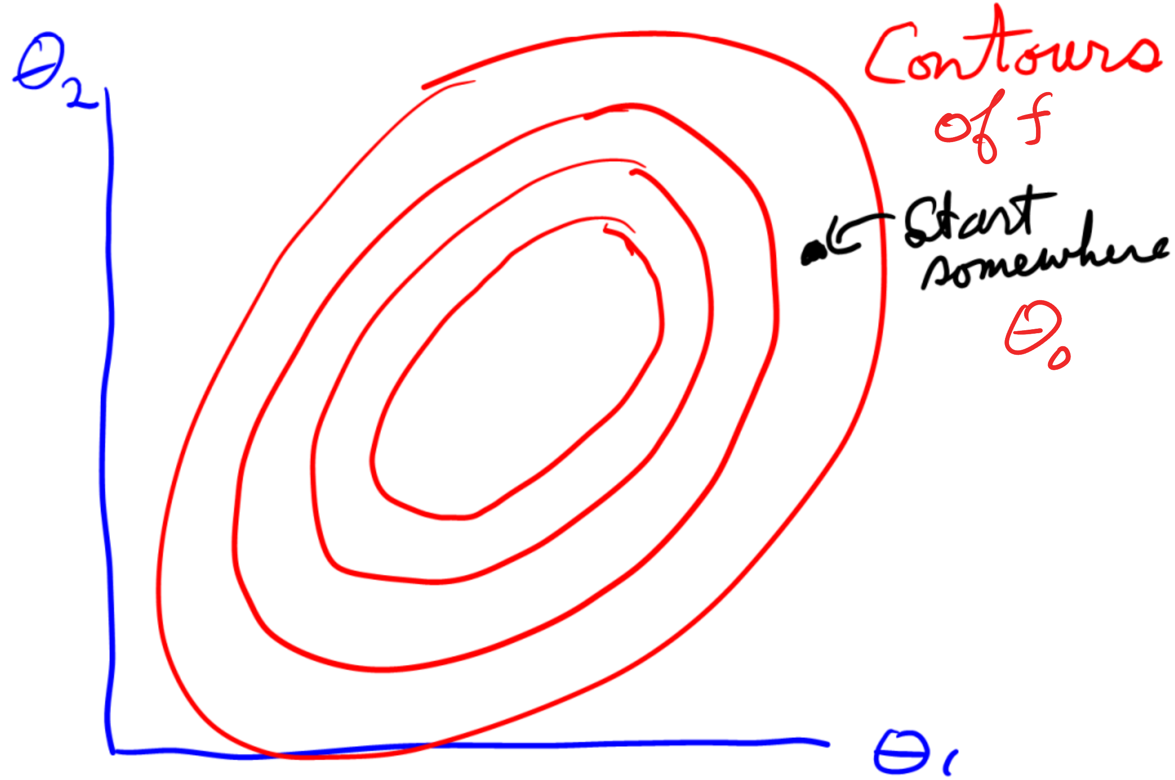
creates a sequence $x_T, \ldots, x_t, x_{t+1}, \ldots x_{\infty}$ that is $\sim f$ but not i.i.d.

Questions:

- How to make this work?
- When will it work?

One way to make it work:

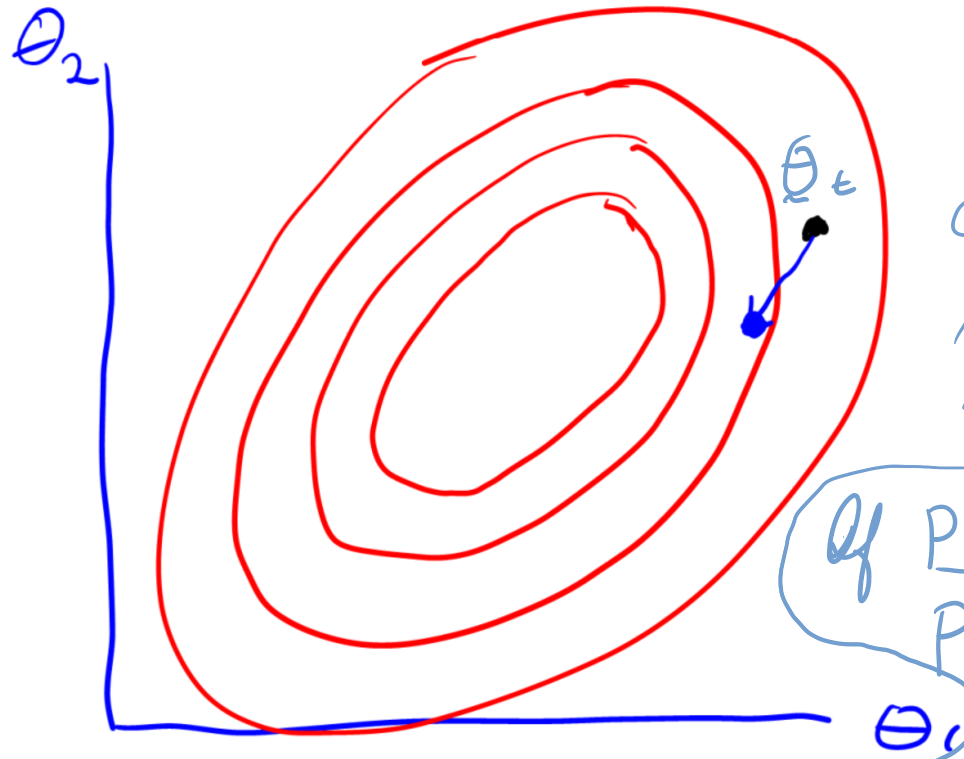
Metropolis-Hastings algorithm.



Using $\underset{\sim}{\theta}$ instead of $\underset{\sim}{x}$ but same idea.

Metropolis-~~Hastings~~ algorithm

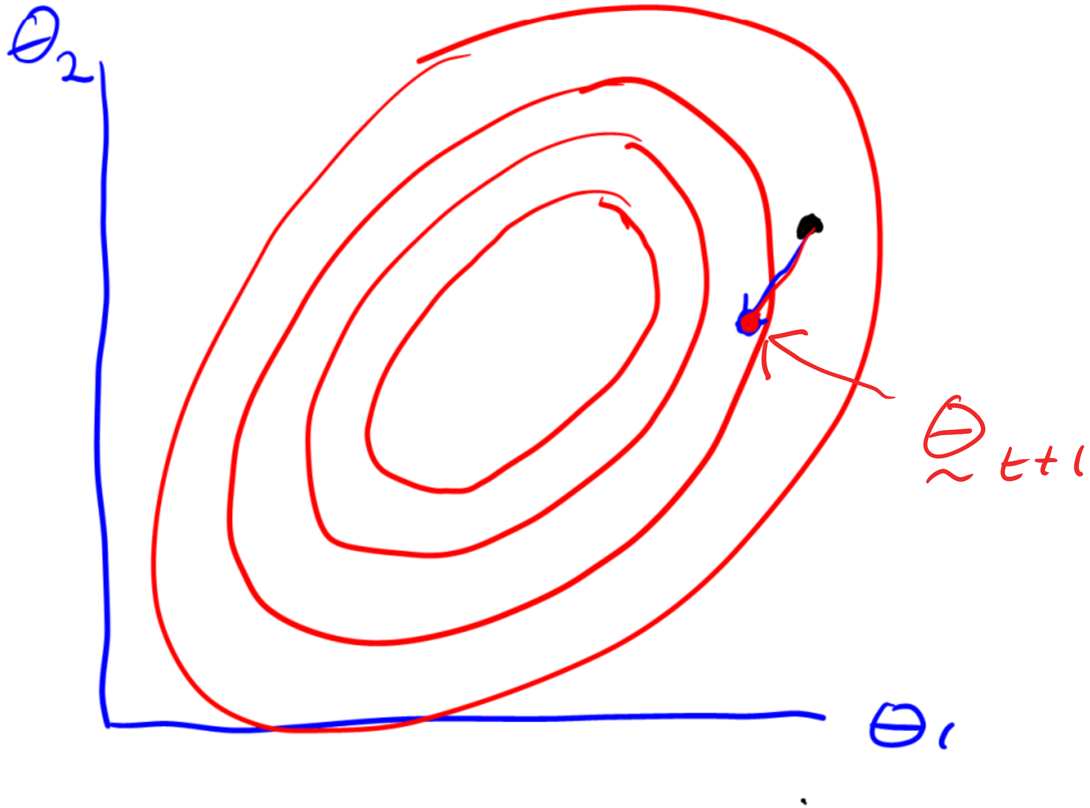
actually this is the special case.



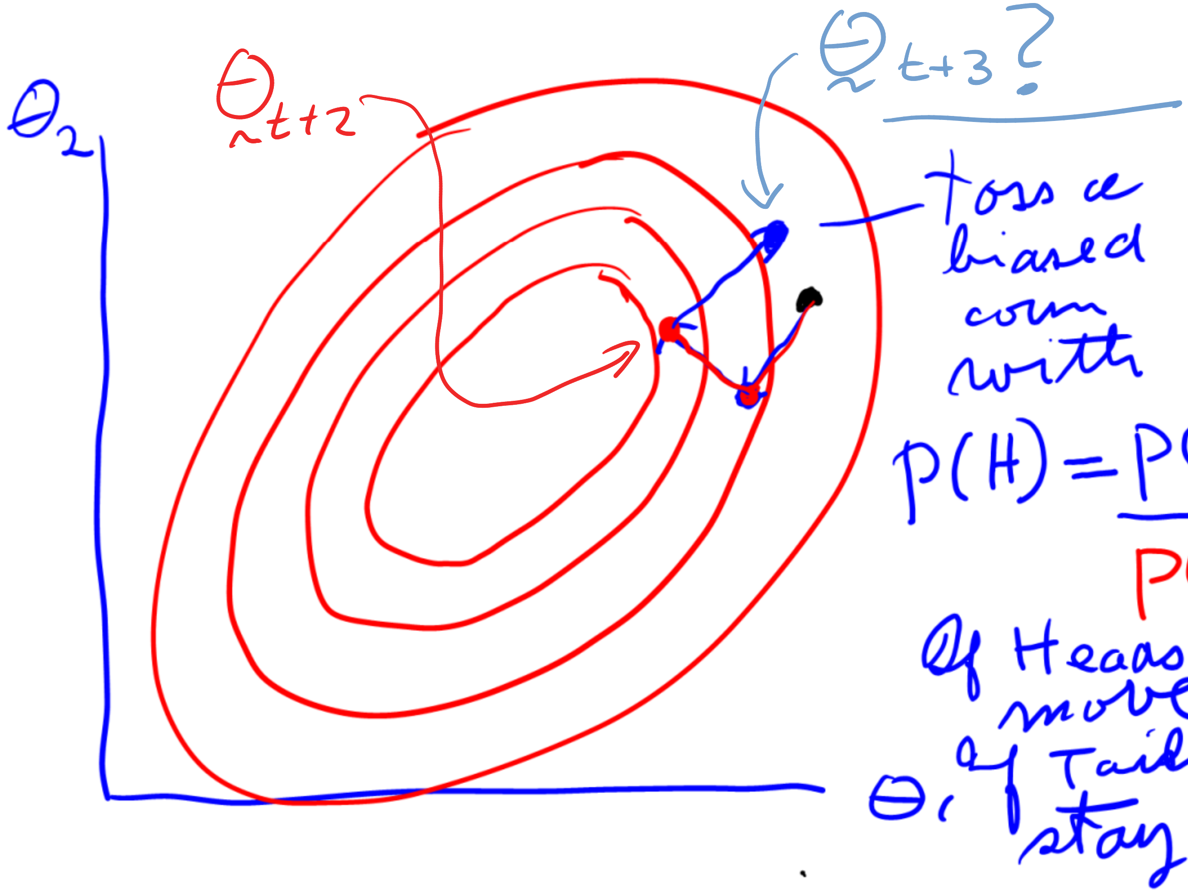
Choose a random step.

If $\frac{P(\bullet)}{P(\bullet)} > 1$ then

$\theta_{t+1} = \bullet$



θ_2
θ_1
$\underset{\sim}{\theta}_{t+1}$

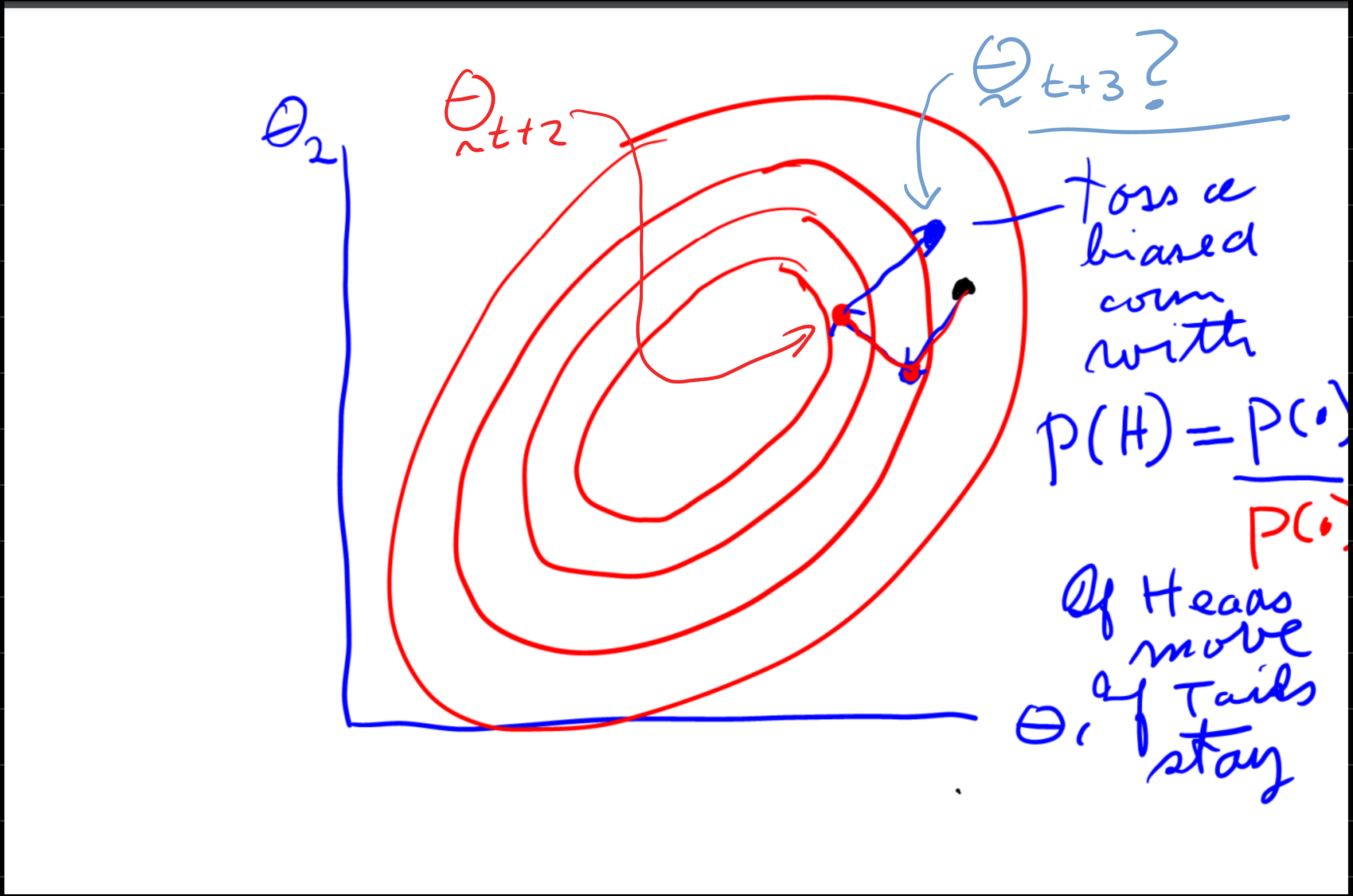


toss a biased coin with

$$P(H) = \frac{P(\cdot)}{P(\cdot)}$$

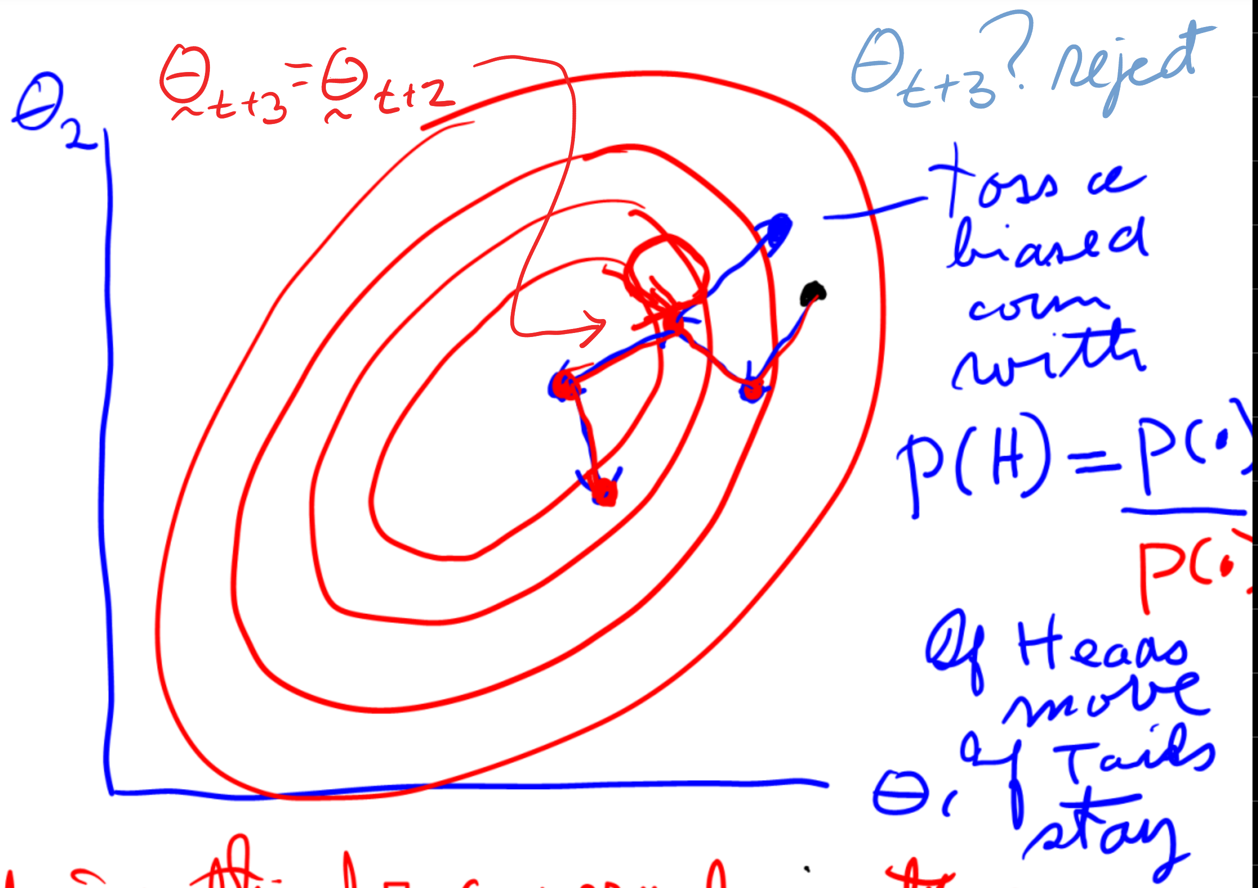
If Heads move
If Tails stay

This is a very big deal! Why?



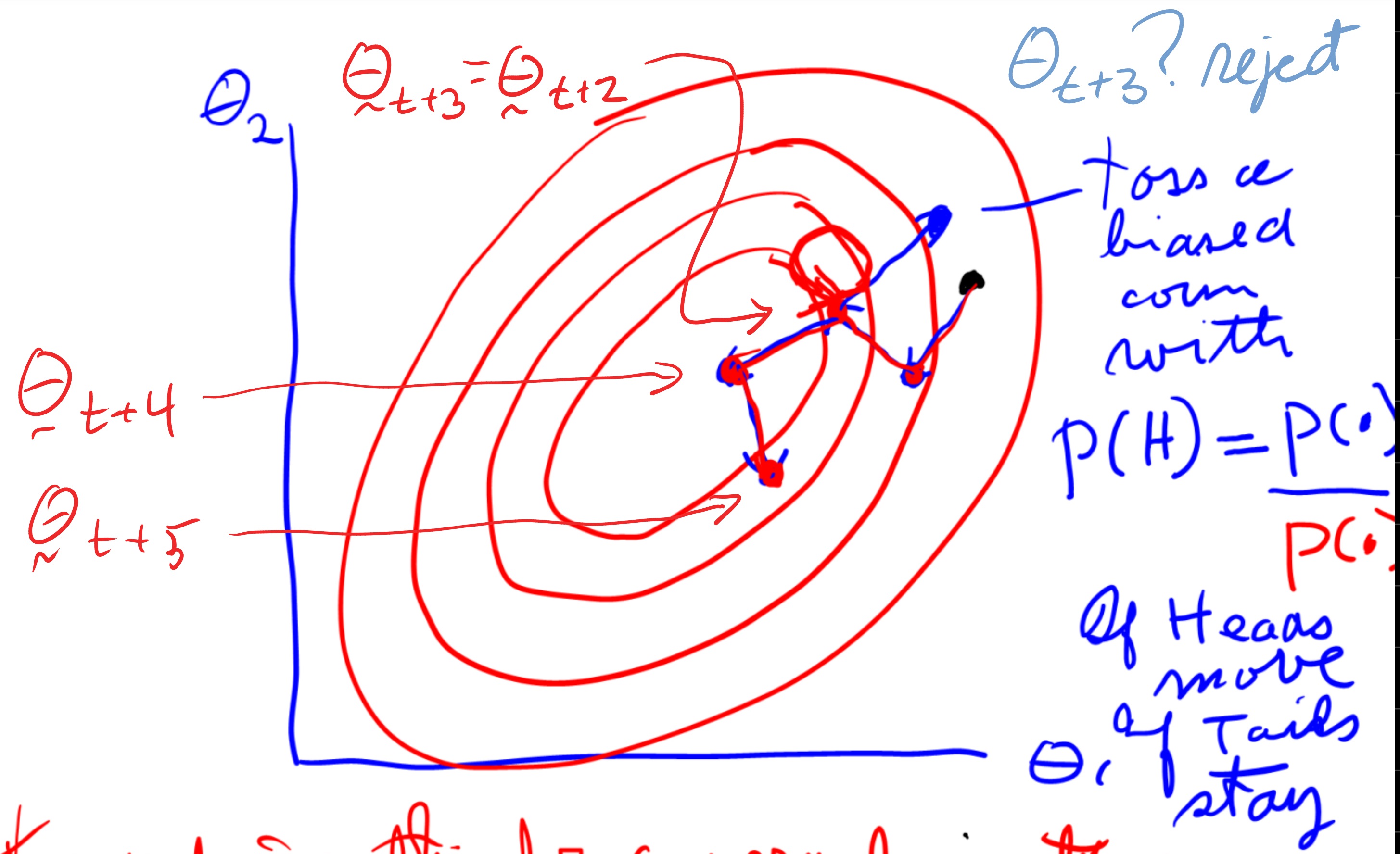
This is a very big deal!

We don't need p itself!! Any $f \propto p$ will do.



Keep doing this for a very long time

Being able to use $f \propto p$ was revolutionary for Bayesian stats.



Keep doing this for a very long time

We get a random chain.

$x_0, x_1, x_2, \ldots, x_T, \ldots, x_t, x_{t+1}, \ldots, x_N$

Problems: (1) early part influenced by choice of x_0

(2) not i.i.d.

Some parts of f might have been ignored.

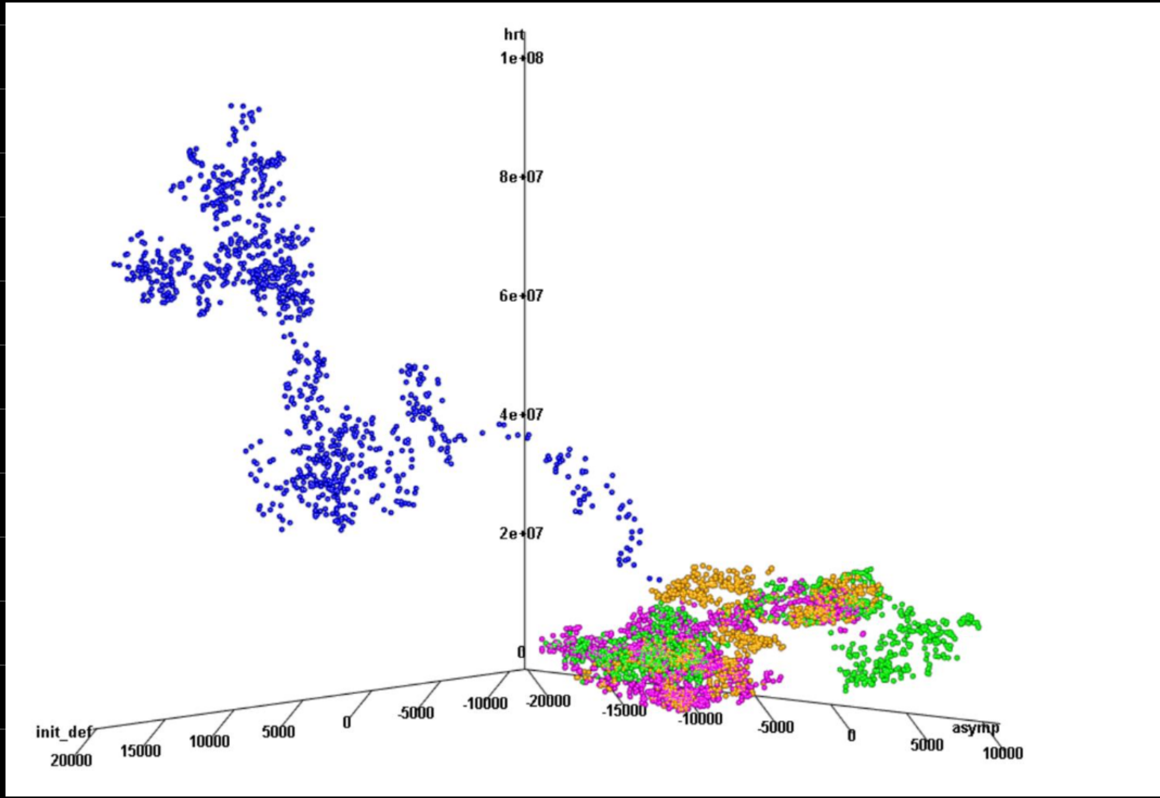
Ideas? : What could we do?

(1) Throw out early part? "Burn in"

(2) Thin the chain: Run longer & keep every 10^kth $\{X_t\}$

③ Create many chains with different starting values and see if they converge.

Diverging chains :

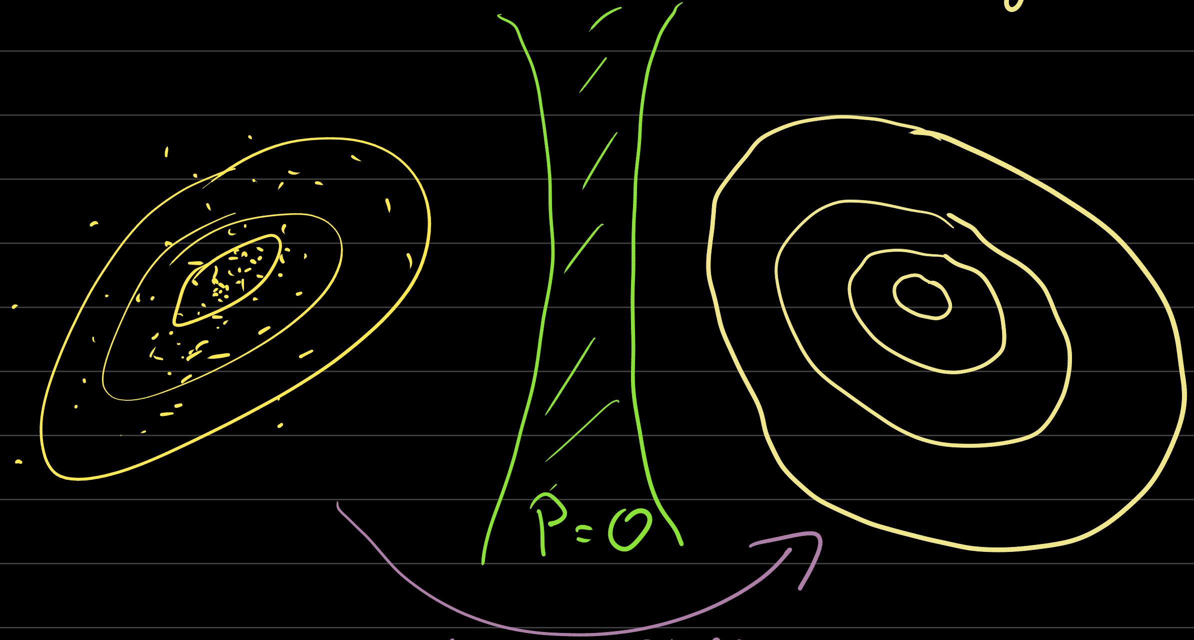


But first:

- When will MCMC utterly fail?
  - Travelling around a mountain range
  - You want to step on every part in a way that is proportional to altitude (f)

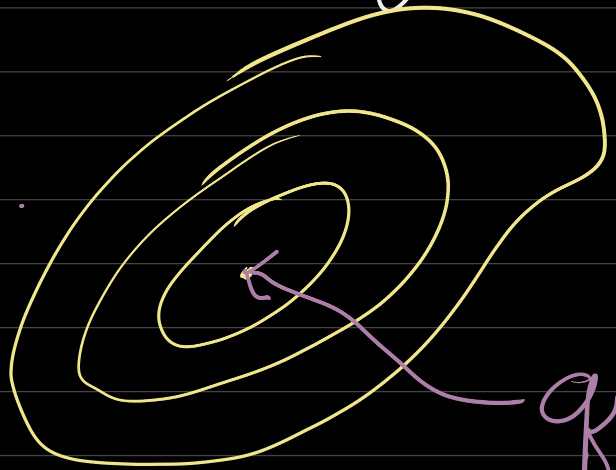
What can go wrong?

1. Parts of range separated by uncrossable chasm (region of 0 cond'l prob.)



MCMC will never reach peaks where it did not start.

② Peaks you can't get off of.



$$q(x \mid x) = 1$$

Theory of MCMC addresses this:

Discrete finite sample space case

$$X_0, X_1, \ldots, X_t, X_{t+1}, \ldots$$

Possible states

with $X_t \in \mathcal{S}$, state space $\mathcal{S} = \{1, \ldots, n\}$

So $q(x_{t+1} \mid x_t)$ can be represented in a <u>transition probability matrix</u>

Markov because ...?

$$P = [p_{ij}], \quad p_{ij} = \Pr(X_{t+1} = j \mid X_t = i)$$

E.g. $S = \{1, 2, 3\}$

rows sum to 1

$$P = \begin{array}{cc} \begin{bmatrix} .1 & .5 & .4 \\ .2 & .0 & .8 \\ .3 & .0 & .7 \end{bmatrix} & \begin{array}{c} \} \ 1 \\ \} \ 1 \\ \} \ 1 \end{array} \end{array}$$

ith row is conditional probs for $X_{t+1} \mid X_t = j$

Suppose $x_0 = 2$

If X_t has pmf $\underset{\sim}{\pi}_t$ (vector of probabilities)

then X_{t+1} has pmf $\pi_t' P$ (work it out)

X_{t+2} has pmf $\pi_t' P^2$

$\vdots$

X_{t+n} has pmf $\pi_t' P^n$

Properties:

$$P \underset{\sim}{1} = \underset{\sim}{1}$$

row sums = 1

$\underset{\sim}{1}$ is a right eigenvector belonging to eigenvalue 1

Definition: P is irreducible

- iff any state can eventually be reached from any other state
- iff $\exists\, m \ni$ all elements of P^m are positive.

Example:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & .5 & .5 \\ 0 & .5 & .5 \end{bmatrix}$$

not irreducible

"is reducible"

Less obvious: $P = \begin{bmatrix} 1 & 0 & 0 \\ .5 & 0 & .5 \\ 0 & .5 & .5 \end{bmatrix}$

is not irreducible. $\{1\}$ is an absorbing state.