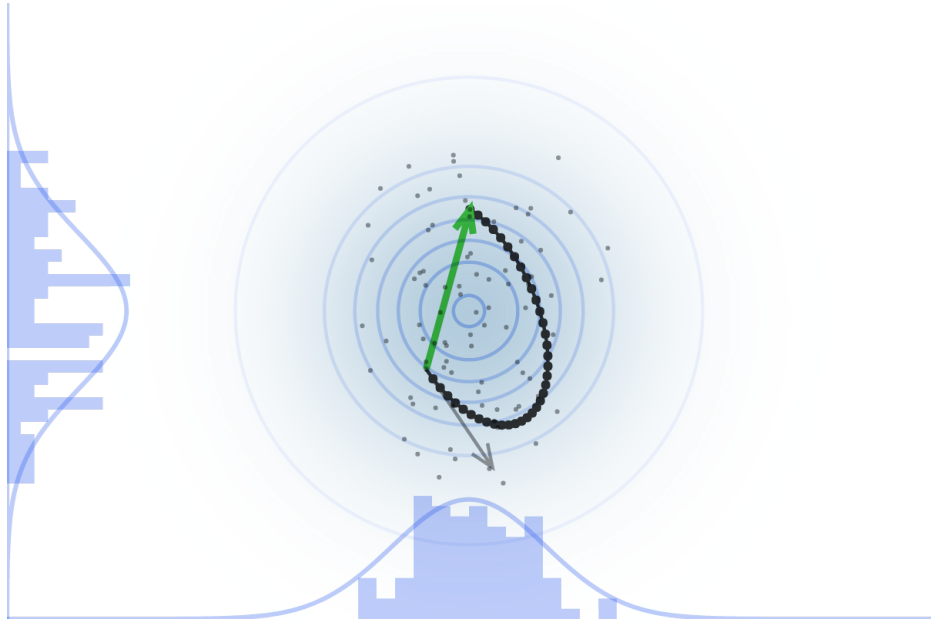
Markov Chain Monte Carlo

Mike Slawinski

March 22, 2018



Agenda

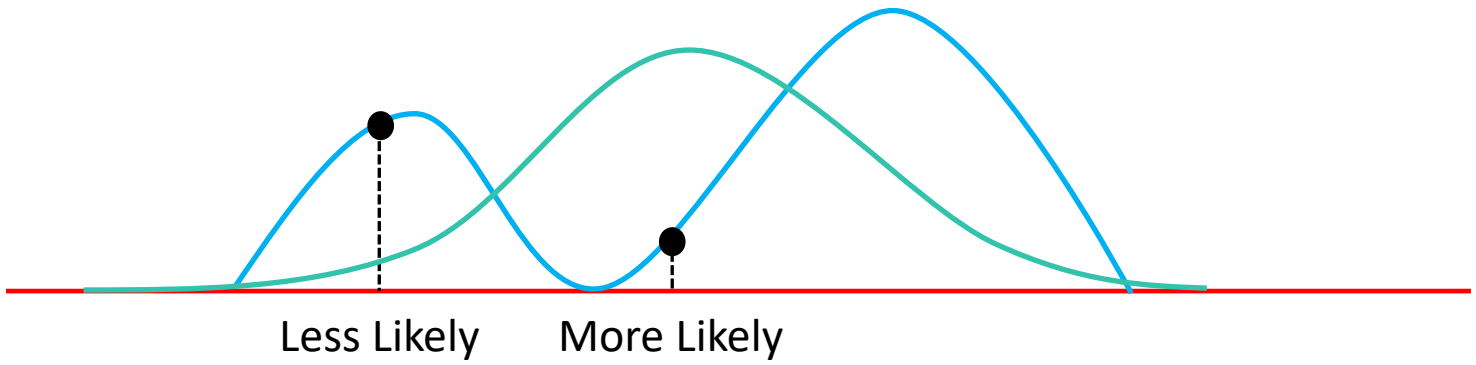
1. Motivation – why do we need to sample?
2. MCMC Example – Metropolis Hastings
3. Measure in High Dimensional Space
4. Hamiltonian Monte Carlo

Motivation

In order to do statistics, we must integrate.

Examples:

- Simple Averaging – consider a random variable X distributed as $p(x)$:
  - $\mu = E[x] = \int xp(x)dx$
  - $\sigma^2 = E[x^2] = \int x^2 p(x)dx$
- Function Averaging – consider a sample space Ω and a function $f: \Omega \to \mathbb{R}$
  - $\mu_f = E[f(x)] = \int_\Omega f(x)p(x)dx$



Motivation – Bayesian Modeling

Understand models by understanding the distributions of the model parameters

Compute the mean of the posterior

$$\mu_\theta = \int \theta p(\theta|x) d\theta = \mathbb{E}_{p(\theta|x)}[\theta]$$

Compute the posterior predictive distribution

$$p(\tilde{x}|\boldsymbol{X}) = \int_\theta p(\tilde{x}|\theta, \boldsymbol{X}) p(\theta|\boldsymbol{X}, \alpha) d\theta = \mathbb{E}_{p(\theta|\boldsymbol{X}, \alpha)}[p(\tilde{x}|\theta, \boldsymbol{X})]$$

Compute marginalization

$$p_X(x) = \int_y p_{X,Y}(x, y) dy = \int_y p_{X|Y}(x|y) p_Y(y) dy = \mathbb{E}_Y[p_{X|Y}(x|y)]$$

Integration via Sampling

Question: How do we integrate?

Answer: Law of Large Numbers

Suppose $X_1, \ldots, \mathrm{X_n}$ is a sequence of i.i.d. random variables with mean $\mathbb{E}[\mathrm{X_i}] = \mu$

Define $\bar{X}_n \coloneqq \frac{1}{n}(X_1 + \cdots + X_n)$

For every $\delta > 0$ and $\epsilon > 0$ there exists a natural number n_0 such that for all $n \geq n_0$,
$P\left(\left|\overline{X}_n - \mu\right| \geq \epsilon\right) \leq \delta$

In other words,

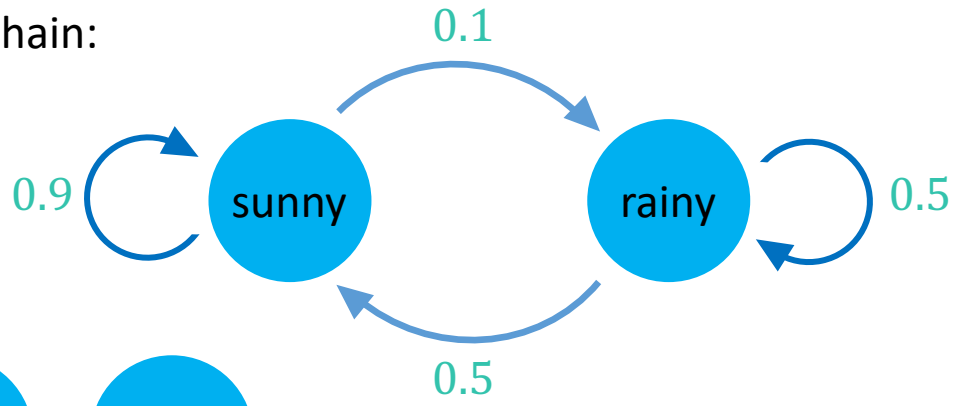
$\lim_{n\to\infty} P\left(\left|\overline{X}_n - \mu\right| \geq \epsilon\right) = 0,$ that is, $\overline{X}_n$ converges in probability to $\mu = \mathbb{E}[X]$

Question: How compute $\mathbb{E}_p[f] = \int f(x)dp(x) = \int f(x)p(x)dx$?

Answer: by sampling i.i.d. $x_1, \ldots, x_n$ from the distribution defined by p and computing $\frac{1}{n}\sum_i f(x_i)$

Markov Chain Monte Carlo – Discrete vs Cont.

Markov Chain:



$\Omega =$ sunny, rainy

$T\left(x^{(i+1)}\middle|x^{(i)}\right) \sim \begin{bmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{bmatrix}$

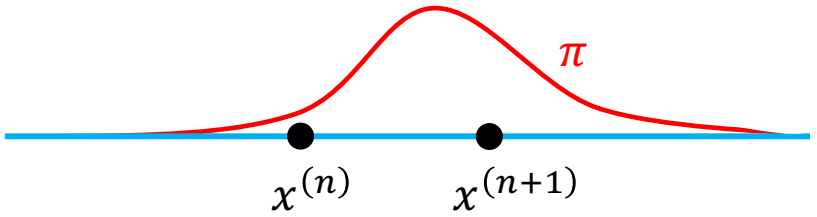
$x^{(n)}$ = sunny/rainy at step n

$\pi^{(n)}$ = Distribution over {sunny,rainy} at step n: $\pi^{(n)} = \pi^{(0)}T^n$

We define the steady state vector as $\pi = \lim_{n\to\infty} \pi^{(n)}$

$= [0.833 \quad 0.167]$

Upshot – moving through this Markov chain ≡ sampling i.i.d. from {sunny,rainy} by flipping the coin $[0.833 \;\; 0.167]$



$\Omega = \mathbb{R}$

$K\left(x^{(i+1)}\middle|x^{(i)}\right)$ - Transition kernel

Continuous Analogue of T,

π = steady state distribution

proportional to target distribution

Upshot – Moving through this Markov chain = sampling i.i.d. from Ω according to the density π

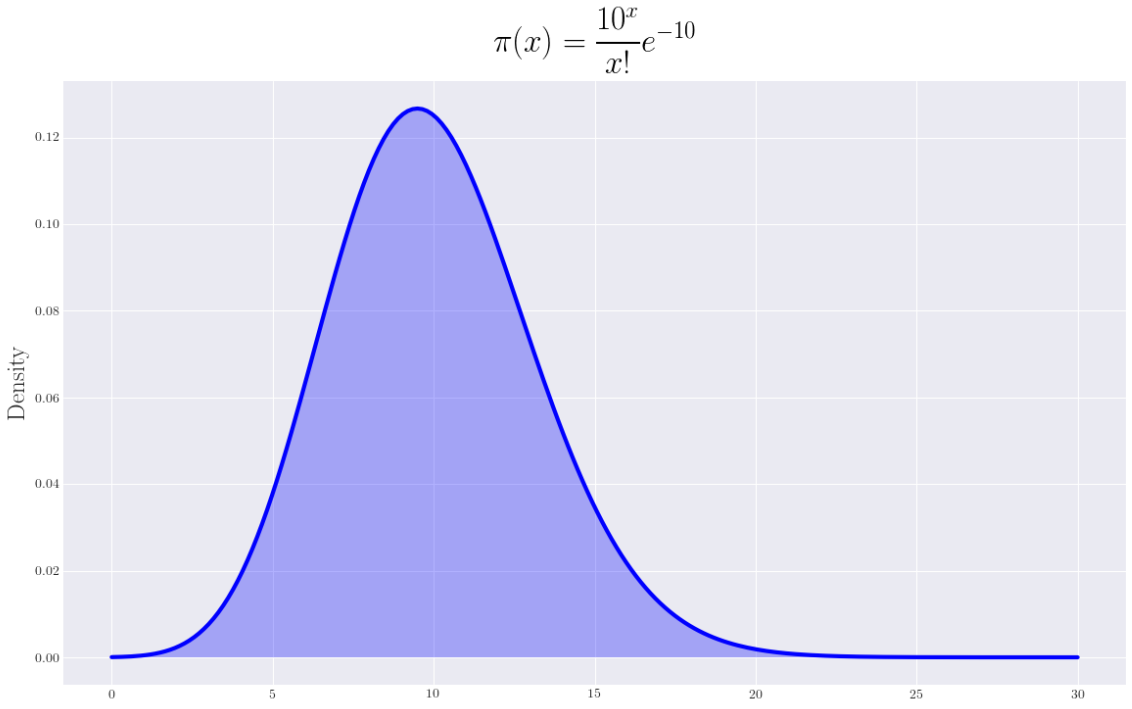
Markov Chain Monte Carlo

Markov chain on the real line

$$K(x_{i+1} = x_i - 1 | x_i) = \begin{cases} \frac{x_i}{20} & if\ x_i \leq 9 \\ \frac{1}{2} & if\ x_i > 9 \end{cases}$$

$$K(x_{i+1} = x_i | x_i) = \begin{cases} \frac{10 - x_i}{20} & if\ x_i \leq 9 \\ \frac{x_i - 9}{2(x_i + 1)} & if\ x_i > 9 \end{cases}$$

$$K(x_{i+1} = x_i + 1 | x_i) = \begin{cases} \frac{1}{2} & if\ x_i \leq 9 \\ \frac{x_i - 9}{5(x_i + 1)} & if\ x_i > 9 \end{cases}$$



Metropolis-Hastings Kernel

Goal: Construct a Markov Chain whose stable distribution is equal to target distribution

$$K_{MH}\left(x^{(i+1)}\middle|x^{(i)}\right) = q\left(x^{(i+1)}\middle|x^{(i)}\right)A\left(x^{(i)}, x^{(i+1)}\right) + \delta_{x^{(i)}}\left(x^{(i+1)}\right)r\left(x^{(i)}\right)$$

q – proposal distribution: Given our current position of $x^{(i)}$, propose a possible next step $x^{(i+1)}$

A – acceptance distribution: Given our current position $x^{(i)}$ and our possible next step $x^{(i+1)}$, what is the probability of actually moving to $x^{(i+1)}$?

δ – Dirac delta $\left(= 1 \text{ if } \mathrm{x}^{(\mathrm{i})} = \mathrm{x}^{(\mathrm{i}+1)}, \text{and zero otherwise}\right)$

$r\left(x^{(i)}\right) = \int_{\mathcal{X}} q\left(x^{*}\middle|x^{(i)}\right)\left(1 - A\left(x^{(i)}, x^{*}\right)\right)dx^{*}$, the probability of rejection

Assume $x^{(i)} = x^{*}$. The kernel has this form because the transition $x^{*} \rightarrow x^{*}$ can occur as a result of a rejected proposal or accepting the proposal $x^{(i+1)} = x^{*}$.

Convergence Criteria

Goal: Construct a Markov Chain whose stable distribution is equal to target distribution

Markov chain converges to a unique limiting distribution π if it is

1. Irreducible
2. Ergodic
    a) Aperiodic
    b) Positive Recurrent

b) Periodic – A state i has period k if any return to state i must occur in multiples of k time steps.
A state is *aperiodic* if $k = 1$.

a) Positive Recurrent State: State i is *positive recurrent* if the expected return time to state i is finite.
A Markov chain is positive recurrent if all of its states are positive recurrent.

We can guarantee that the Markov chain converges to the target distribution if the kernel K_{MH} satisfies the detailed balance equation

Convergence Criteria

Balance Condition:

$$\pi\left(x^{(i)}\right)K_{MH}\left(x^{(i+1)}\middle|x^{(i)}\right) = \pi\left(x^{(i+1)}\right)K_{MH}\left(x^{(i)}\middle|x^{(i+1)}\right)$$

Discrete detailed balance condition:

$$\pi\left(x^{(i)}\right)T\left(x^{(i-1)}\middle|x^{(i)}\right) = \pi\left(x^{(i-1)}\right)T\left(x^{(i)}\middle|x^{(i-1)}\right)$$

Summing both sides yields

$$\pi\left(x^{(i)}\right) = \sum_{x^{(i-1)}} \pi\left(x^{(i-1)}\right)T\left(x^{(i)}\middle|x^{(i-1)}\right)$$

Continuous version:

$$\int \pi\left(x^{(i)}\right)K\left(x^{(i+1)}\middle|x^{(i)}\right)dx^{(i)} = \pi\left(x^{(i+1)}\right)$$

Intuition: this condition implies that if we generate an ensemble of samples from the target distribution and applied the transition, then we would get a new ensemble that was still distributed according to the target distribution

$$\pi_i T_{ij} = \pi_j T_{ji} \; \forall i, j$$

$\pi_i T_{ij}$ represents the amount of probability that flows down edge $i \to j$ in one time step. The amount of probability flowing from i to j equals the amount that flows from j to i. There is therefore no net flux of probability along the edge $i \leftrightarrow j$. This is equivalent to the density over the nodes being fixed.

MCMC - Metropolis Hastings

Balance Condition Holds

$$p\left(x^{(i)}\right)K_{MH}\left(x^{(i+1)}\middle|x^{(i)}\right) = p\left(x^{(i+1)}\right)K_{MH}\left(x^{(i)}\middle|x^{(i+1)}\right)$$

for kernel

$$K_{MH}\left(x^{(i+1)}\middle|x^{(i)}\right) = q\left(x^{(i+1)}\middle|x^{(i)}\right)A\left(x^{(i)}, x^{(i+1)}\right) + \delta_{x^{(i)}}\left(x^{(i+1)}\right)r\left(x^{(i)}\right)$$

by setting

$$\mathcal{A}\left(x^{(i)}, x^{\star}\right) = \min\left\{1, \frac{p(x^{\star})q\left(x^{(i)}\right)}{p\left(x^{(i)}\right)q(x^{\star})}\right\}$$

Biased towards the mode for q symmetric

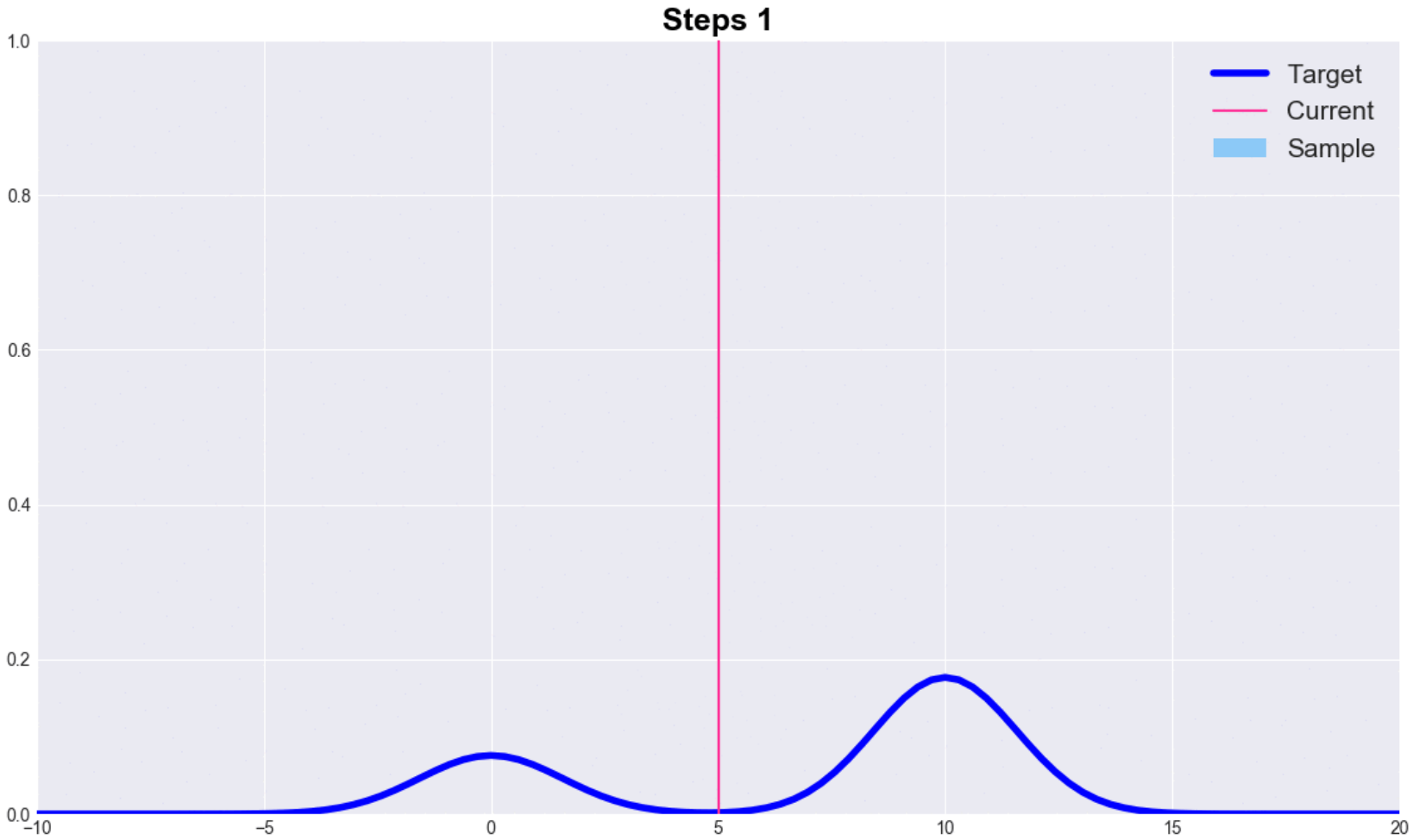
Algorithm

Metropolis-Hastings(x_0, P, Q)

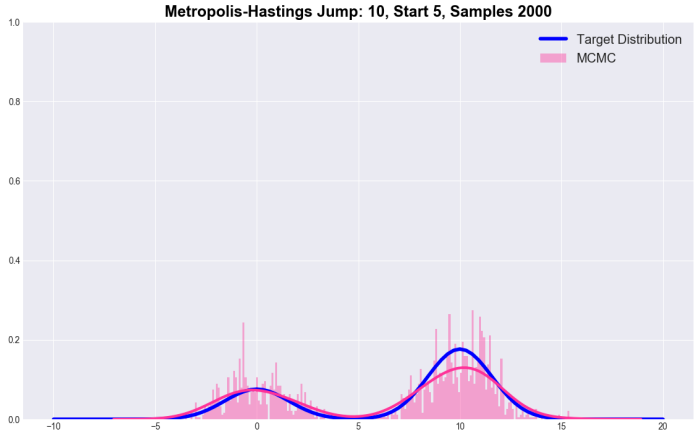
For $k = 0, .., N-1$ do:

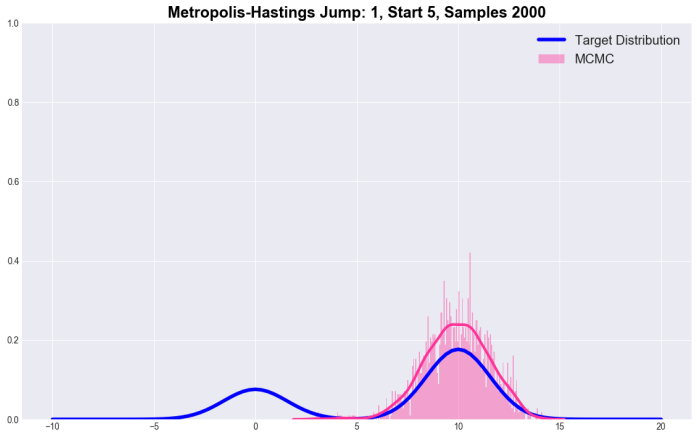
1. Sample x' from $Q(x'|x_k)$.
2. With *acceptance probability* $\alpha(x', x_k) = \min\left(1, \frac{P(x')Q(x_k|x')}{P(x_k)Q(x'|x_k)}\right)$, set $x_{k+1} \leftarrow x'$.
3. Otherwise, set $x_{k+1} \leftarrow x_k$.

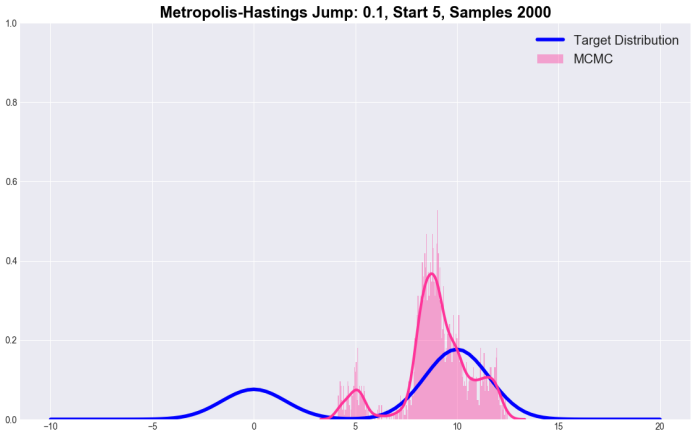
Metropolis - Hastings

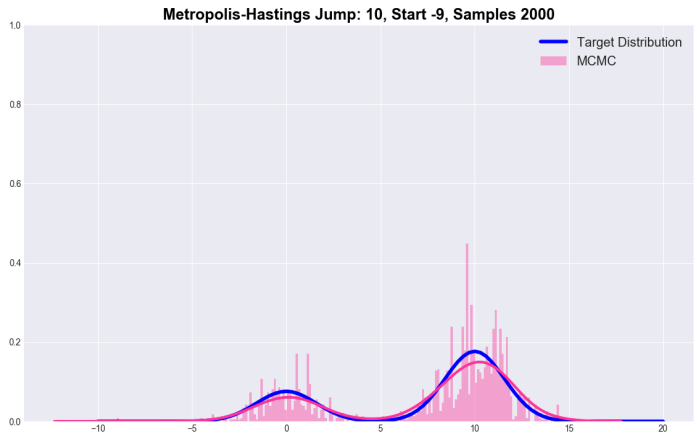


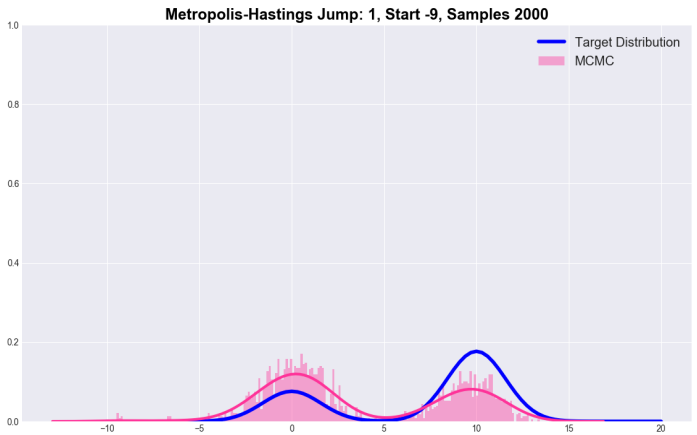
Metropolis – Hastings : Parameter Grid

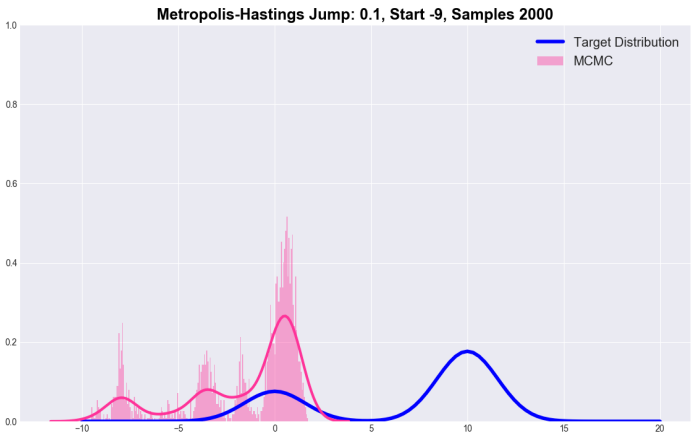












MCMC - Metropolis Hastings in High Dim

Algorithms like MH tend to fail in high dimensions.

Why?

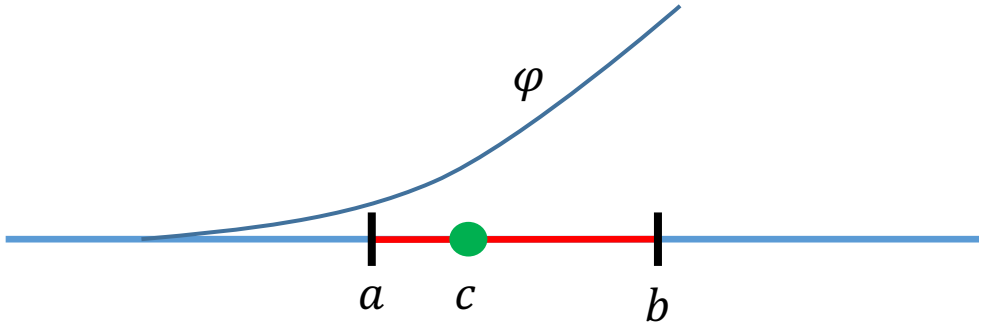
Because $q\left(x^{(i+1)}\middle|x^{(i)}\right)$ for q naïve mistakes density for mass

Measure Theory – Density vs Mass

Let X be a set and Σ a $\sigma-$ algebra over X. A function $m: \Sigma \to \mathbb{R} \cup \{+\infty, -\infty\}$ is a measure if it satisfies the following:

- Non-negativity: For all E in Σ: $m(E) \geq 0$
- Null empty set: $m(\emptyset) = 0$
- Countable additivity: $m(\bigcup_{k=1}^{\infty} E_k) = \sum_{k=1}^{\infty} m(E_k)$
- Lebesgue Measure:
  - $m_{\mathcal{L}}([a,b]) = b - a$
- Dirac Measure:
  - $\delta_c([a,b]) = 1$
- φ $-$ defined Measure:
  - $m_{\varphi}([a,b]) = \int_{[a,b]} d\varphi = \int_a^b \varphi dx$
  - P defined by a density φ: $P([a,b]) = \int_{[a,b]} d\varphi$

What is the size of the interval $[a,b]$?

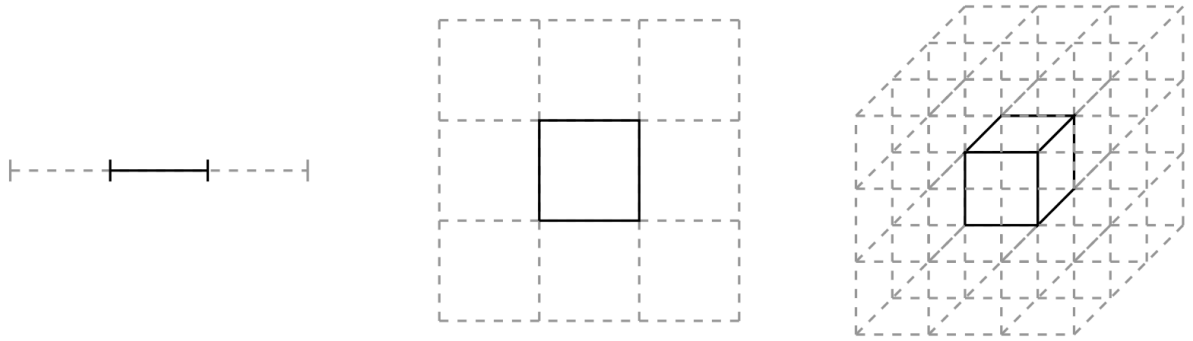


Example:

$$p(\tilde{x}|\boldsymbol{X}) = \int_{\theta} p(\tilde{x}|\theta, \boldsymbol{X}) \overbrace{p(\theta|\boldsymbol{X}, \alpha) d\theta}^{\text{measure}} = \mathbb{E}_{p(\theta|\boldsymbol{X}, \alpha)}[p(\tilde{x}|\theta, \boldsymbol{X})]$$

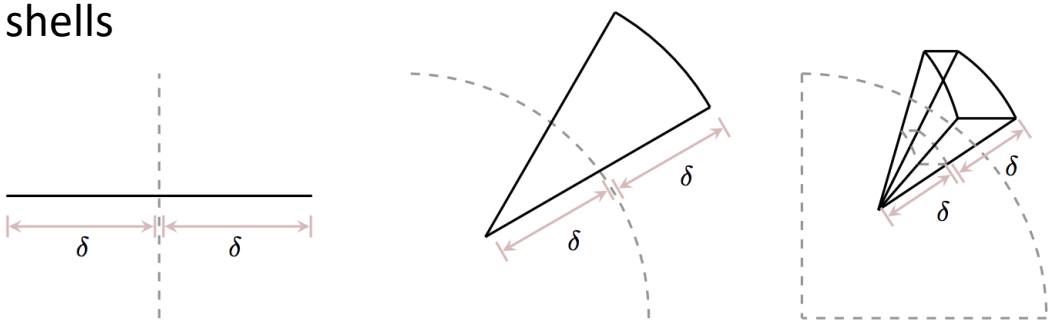
Volumes in High Dimensions

Distribution of volume as a function of dimension:

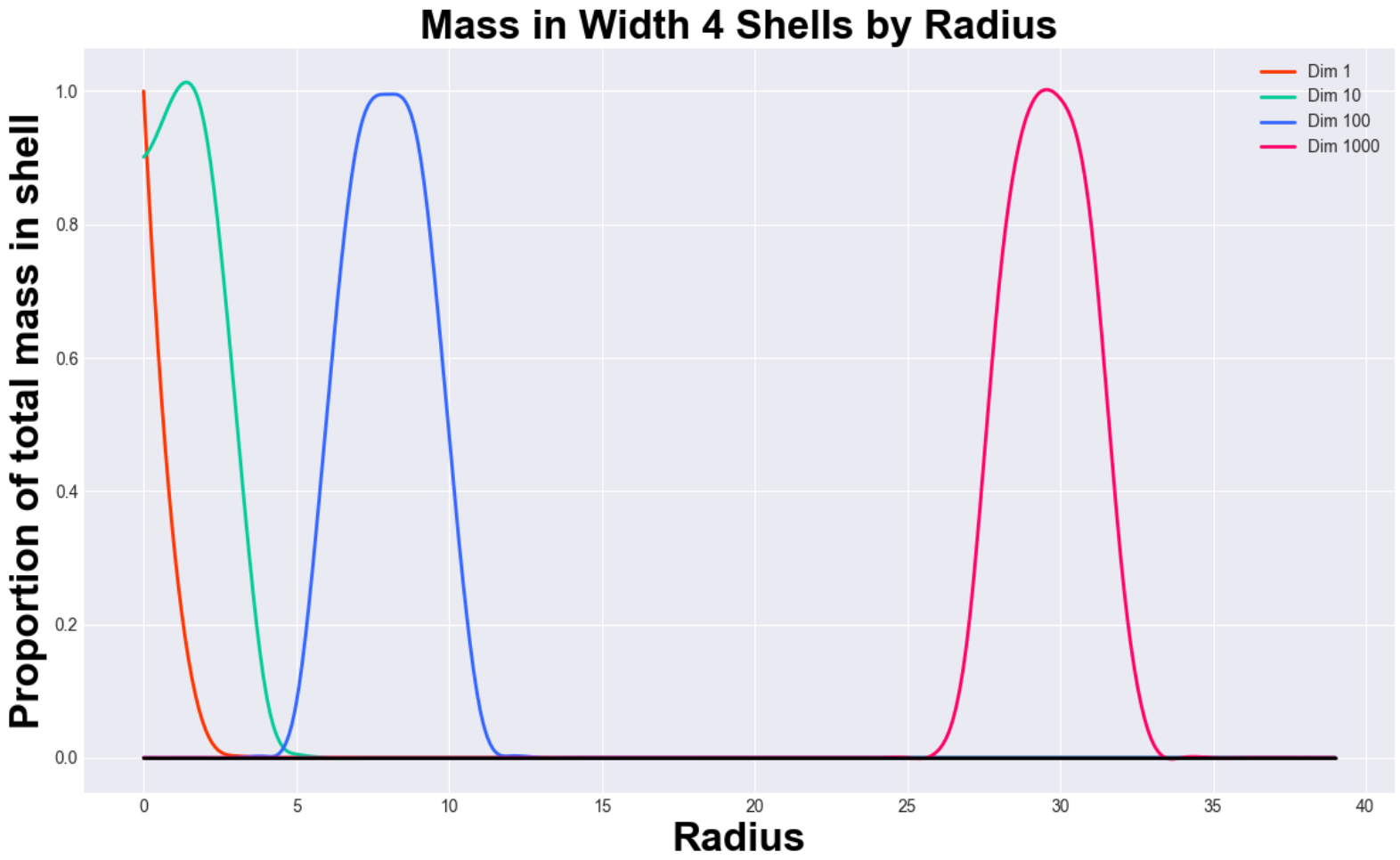


$3^D - 1$ neighboring partitions in D − dimensional space

In terms of concentric spherical shells



The Typical Set: Example



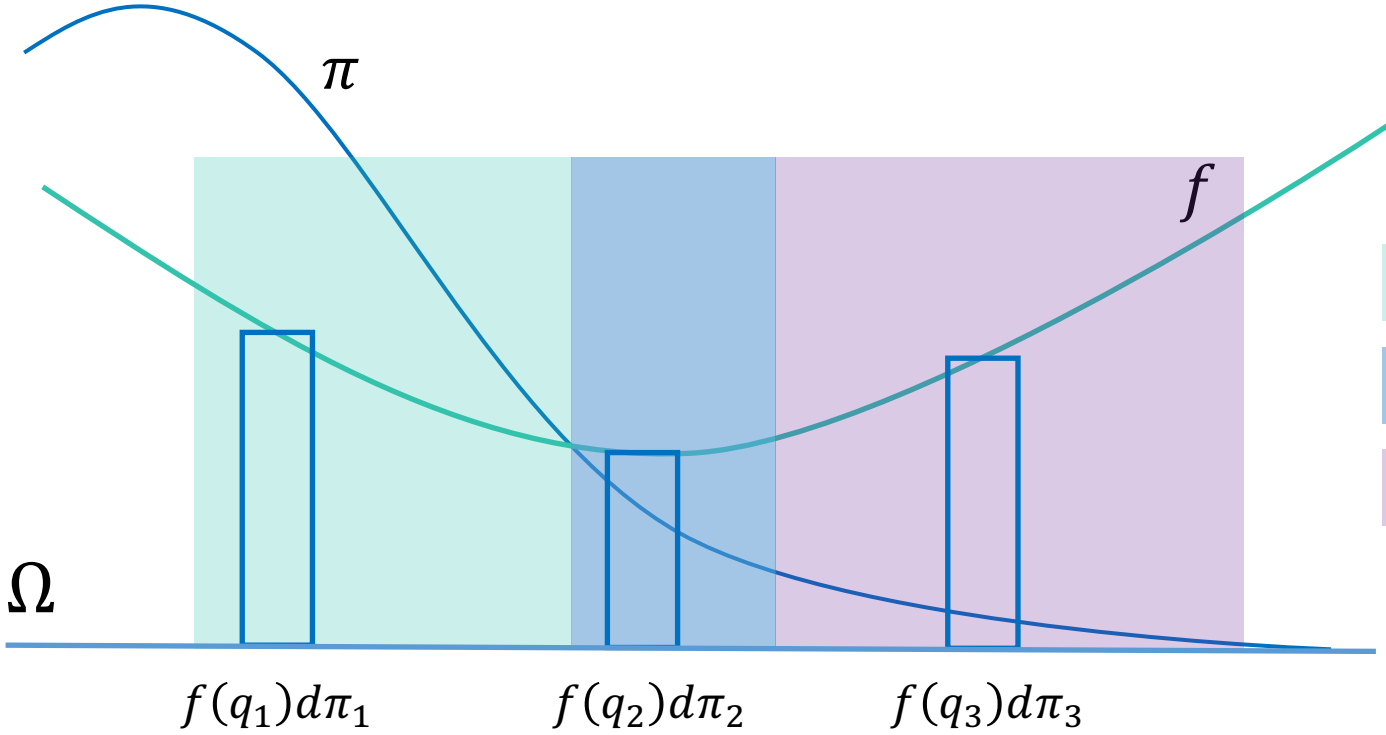
$d = 1$: $99.73\%\ of\ mass\ in\ S(0.1,3)$

$d = 10$: $99.44\%\ of\ mass\ in\ S(1,5)$

$d = 100$: $99.53\%\ of\ mass\ in\ S(8,12)$

$d = 1000$: $99.00\%\ of\ mass\ in\ S(30,34)$

Integration over the Typical Set



Low Probability Mass – High density, Low measure

High Probability Mass – Medium density, Medium measure

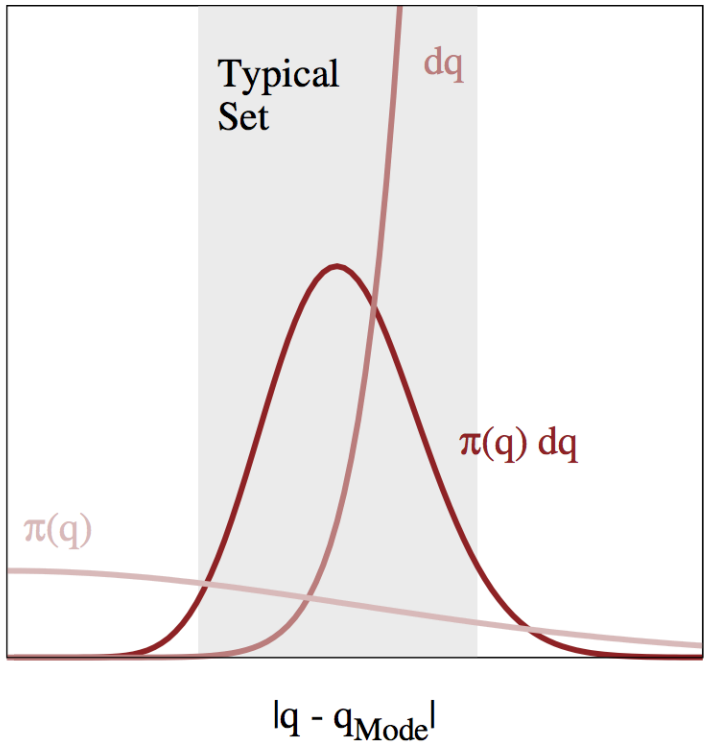
Low Probability Mass – Low density, High measure

Partition the entire sample space Ω

Rectangle Area $= f(q_2)d\pi_2 \gg f(q_1)d\pi_1, f(q_3)d\pi_3$

$$\int_{\Omega} f d\pi \sim \int_{\blacksquare} f d\pi \sim \frac{1}{n}\sum_{q\sim\blacksquare} f(q)$$

The Typical Set



$\boldsymbol{dq}$ = **measure:** curve expresses Lebesgue measure of $\{q^* \; s.t. \; |q^* - q_{mode}| < q\}$ as a function of distance from the mode, i.e., the volume contained the $n-$sphere of radius $|q - q_{Mode}|$.

$\boldsymbol{\pi(q)}$ = **density:** curve expresses probability density defined by π as a function of distance from the mode

$\boldsymbol{\pi(q)dq}$ = **mass:** curve expresses the probability mass as a function of distance from the mode

Note: $\boldsymbol{\pi(q)dq = d\pi(q)}$: measure defined by the density π as a function of distance from the mode

Naïve intuition – simply use calculus to find the modes of the density and use density-based MCMC with small step size to explore each mode independently.

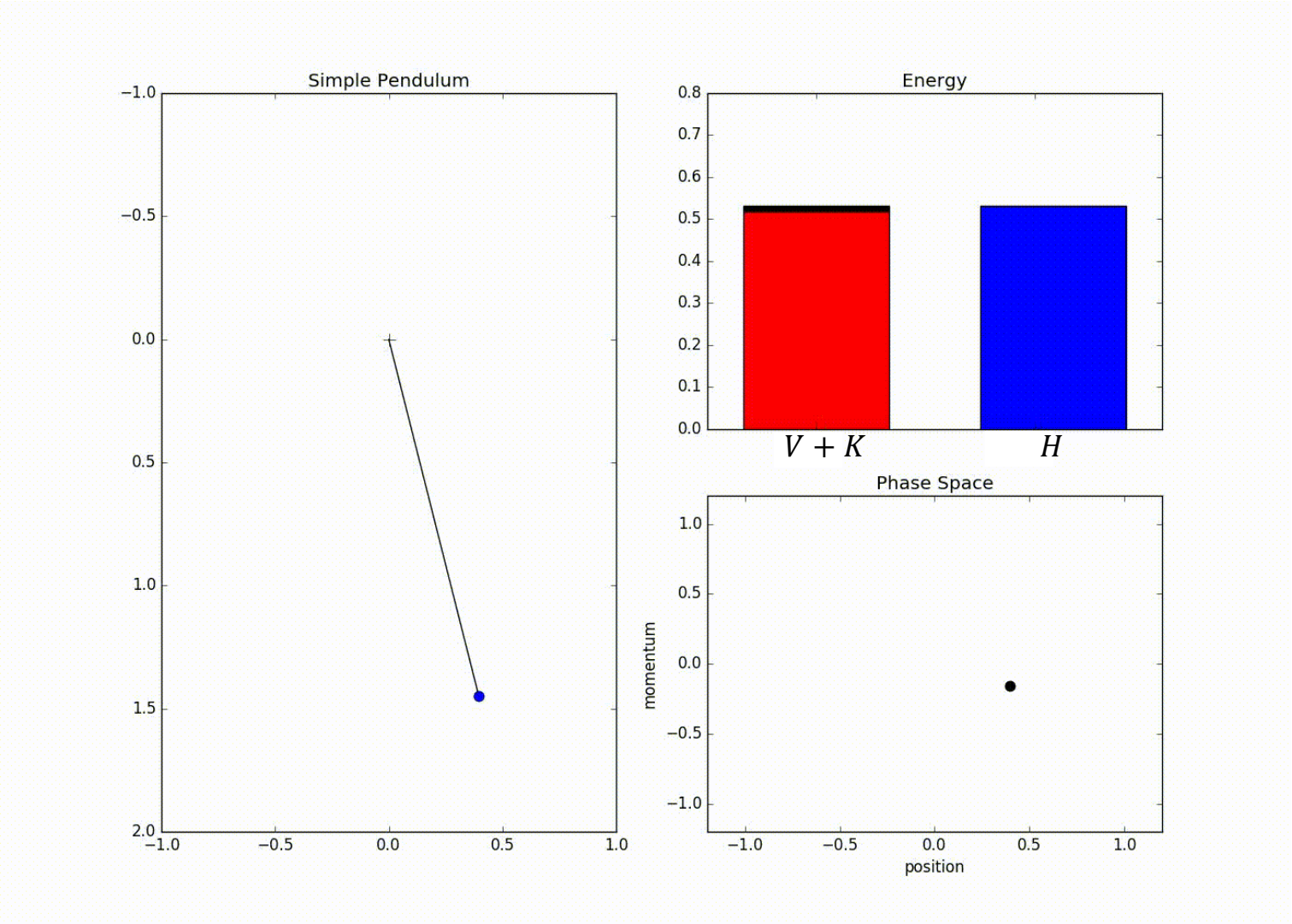
Better Intuition – No need to integrate in regions where $d\pi(q) = \pi(q)dq \ll 0$

Integration over the Typical Set

Question: How can we take advantage of the geometry of our distribution so that our Markov Chain spends the vast majority of its time sampling the typical set?

Answer: Hamiltonian Monte Carlo

Hamiltonian Mechanics – Phase Space



$H = K + V$

K = Kinetic Energy

V = Potential Energy

Black – Kinetic Energy
Red – Potential Energy
Blue – Total Energy (constant)

Hamiltonian Monte Carlo

Hamilton's Equations:

Trajectories obtained by integrating this vector field are energy preserving – Markov chain proposals sit at the ends of these trajectories.

Hamiltonian Monte Carlo – Hamilton's Eqns

Hamilton's Equations:

Idea – describe the dynamics of a system by considering the total energy of the system in terms of canonical coordinates in phase space (q, p): q – spatial coordinates, p – momentum coordinates

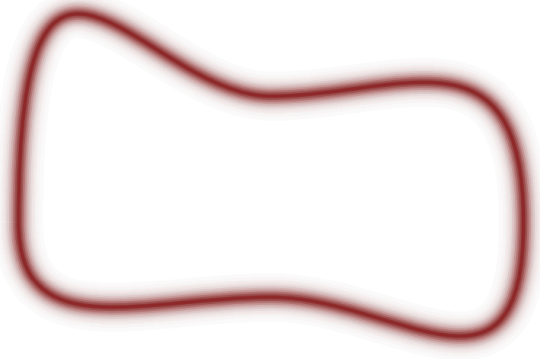
Time evolution of the system described by **Hamilton's Equations:**

$$\frac{dq}{dt} = +\frac{\partial H}{\partial p} = \frac{\partial K}{\partial p}$$

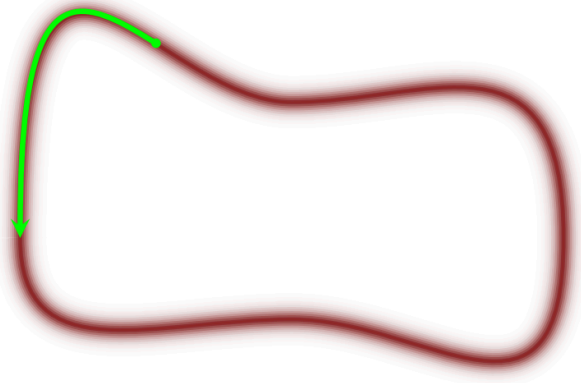
$$\frac{dp}{dt} = -\frac{\partial H}{\partial q} = -\frac{\partial K}{\partial q} - \frac{\partial V}{\partial q}$$

Trajectories obtained by integrating this vector field are energy preserving – Markov chain proposals are sit at the ends of these trajectories.

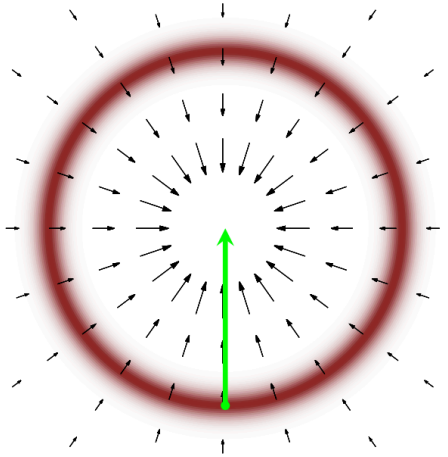
Hamiltonian Monte Carlo – Vector Fields



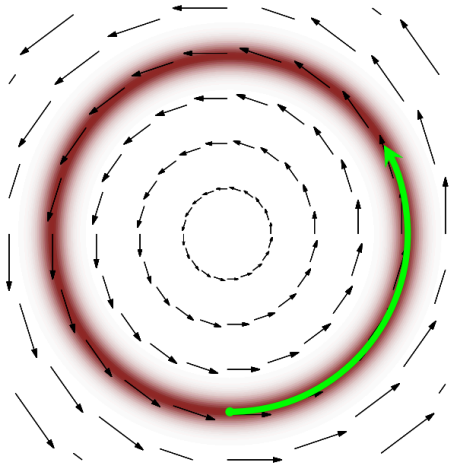
Typical Set



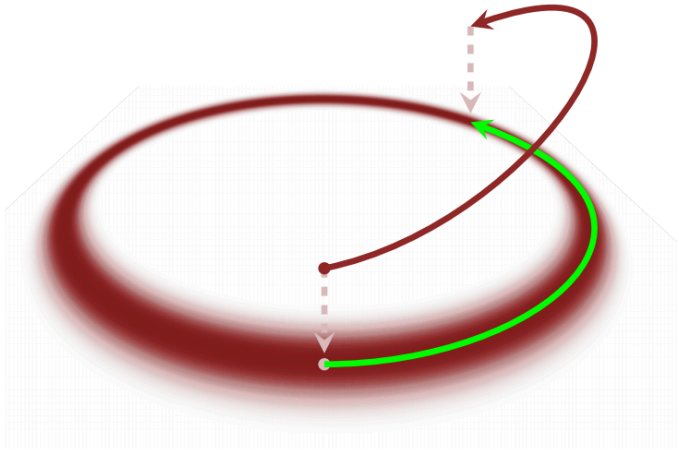
Desired Markov Step Through Typical Set



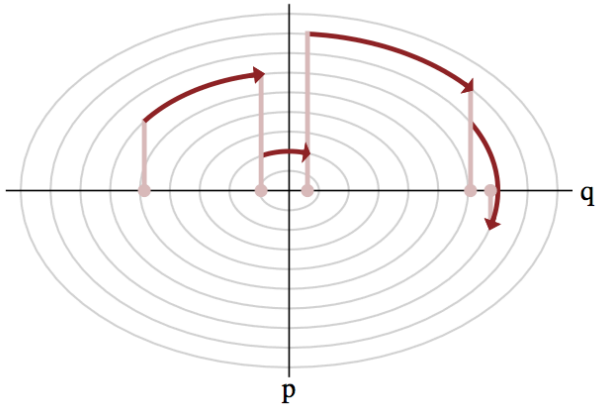
Step based solely on density gradient



Preferred vector field



Phase (position,momentum) space



Energy level sets in phase space

Hamiltonian Monte Carlo – Energies

Kinetic energy as a function of momentum in one dimension takes the form $K(p) = \frac{p^2}{2m}$

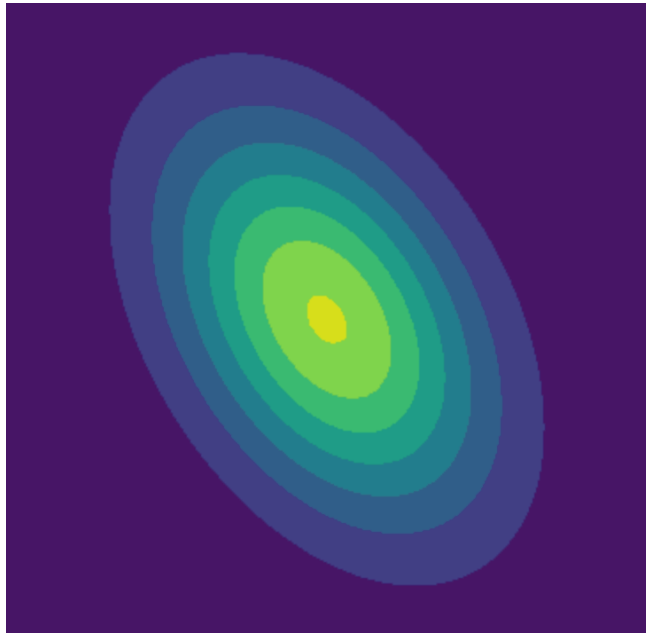
In higher dimensions $K(q,p)$ should look like quadratic form in p: $K \sim p^T A\, p$

Is there a matrix A which encodes the geometry of the distribution?

Set the kinetic energy to $K(q,p) = \frac{1}{2}p^T M^{-1} p + \log|M| + const$
for $M^{-1} = \mathbb{E}[(q-\mu)(q-\mu)^T]$

Equivalently,

$$e^{-K(q,p)} \propto N(p|0,M)$$



Potential energy should be a function of the gradient of the density

Following the skateboarder analogy, we should set the potential energy to $\mathrm{V(q)} = -\log \pi(q)$

Hamiltonian Monte Carlo – Momentum

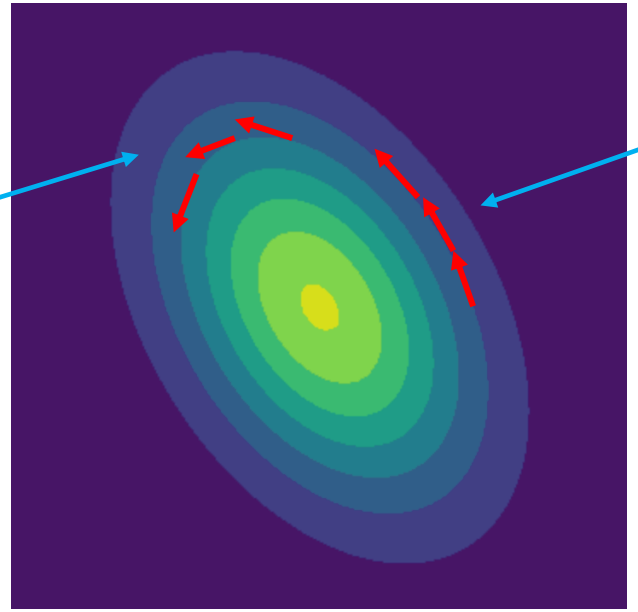
Intuition – when traveling along trajectories in projected space we have two families of states:

Small Changes in Momentum Space

$dist\big(p_i(t), p_i(t+\Delta t)\big)$large

Large Changes in Position Space

$dist\big(q_i(t), q_i(t+\Delta t)\big)$ small



Small Changes in Momentum Space

$dist\big(p_i(t), p_i(t+\Delta t)\big)$small

Large Changes in Position Space

$dist\big(q_i(t), q_i(t+\Delta t)\big)$large

Upshot: The metric in position space should be dual to the metric in momentum space.

Position Space Metric: $dist(q, q') = (q-q')^T \boldsymbol{M} (q-q')$

Momentum Space Metric: $dist(p, p') = (p-p')^T \boldsymbol{M^{-1}} (p-p')$

Hamiltonian Monte Carlo – Canonical Dist

Lift the target distribution $\pi(q)$ to a distribution over $\pi_{Q,P}(q,p) = \pi_{P|Q}(p|q)\pi_Q(q)$ phase space such that

$$\pi(q) = \int \pi_{Q,P}(q,p)dp = \int \pi_{P|Q}(p|q)\pi_P(p)dp$$

Given a closed system with energy E, we can associate a probability distribution P called the canonical distribution

$$P = \frac{1}{Z}\exp\left(-\frac{E}{T}\right)$$

which assigns a probability to each possible state of the system – this is exactly what we want, i.e., a way of constructing trajectories through phase space by via sampling

In our case, the energy E is given by the Hamiltonian $H(q,p) = K(q,p) + V(q)$

This yields $e^{-H(q,p)} = e^{-K(q,p)}e^{-V(q)} = \pi(p|q)\pi(q)$

and we should set

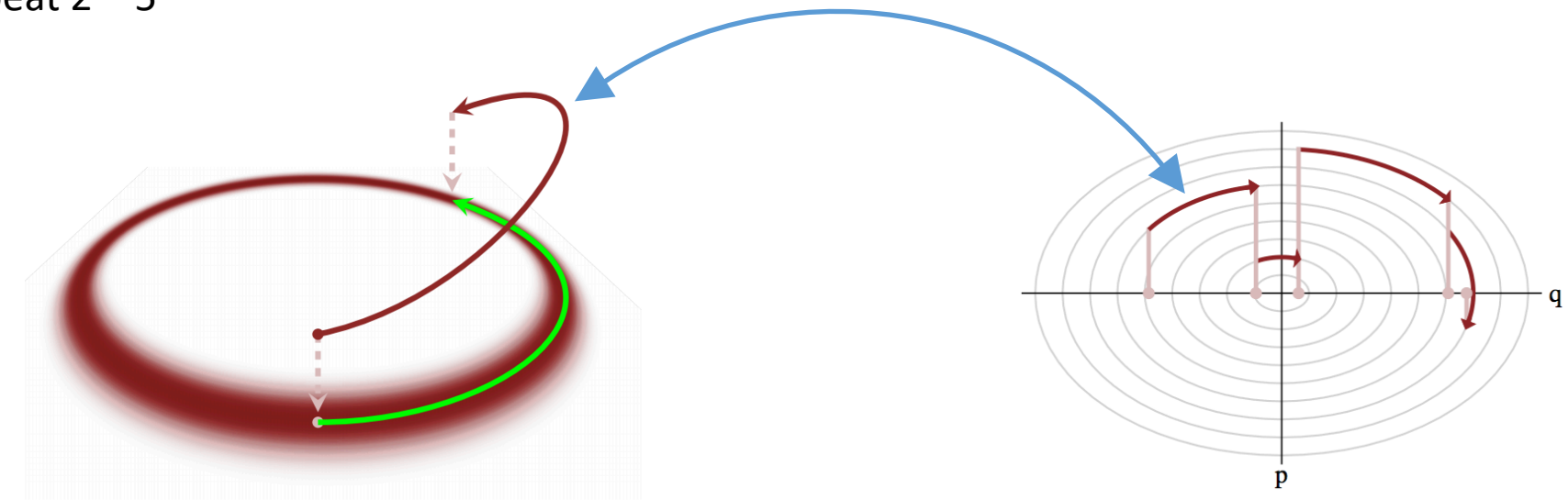
$$\pi(p|q) \coloneqq e^{-K(q,p)} \text{ or } K(q,p) = -\log\pi(p|q) \quad \text{(Prev Slide: } \pi(p|q) = N(p|0,M)\text{)}$$

$$\pi(q) \coloneqq e^{-V(q)} \text{ or } V(q) = -\log\pi(q)$$

Hamiltonian Monte Carlo – Trajectories

Outline Algorithm at High Level:

1. Choose a point in parameter space q_0
2. Randomly sample from the momentum distribution $p_0 \sim \pi(p|q)$
3. Run the energy-preserving trajectory forward to (q^*, p^*)
4. Project (q^*, p^*) to q^*
5. Check acceptance
6. Repeat 2 – 5

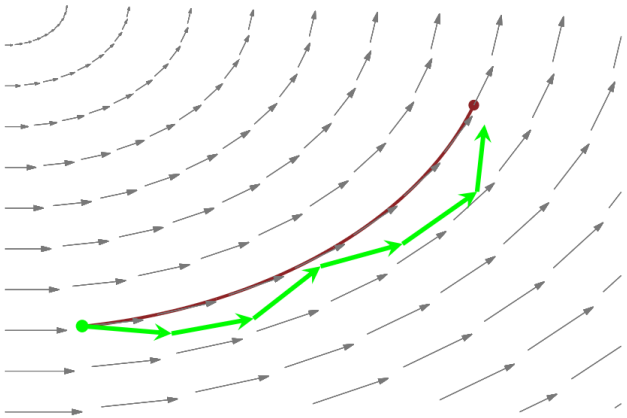


Hamiltonian Monte Carlo – Leap Frog

How do we integrate this vector field to obtain energy-preserving trajectories through phase space?

$q_0 \leftarrow q, p_0 \leftarrow p$
for $1 \leq n \leq \lfloor T/\epsilon \rfloor$ **do**
$p_{n+\frac{1}{2}} \leftarrow p_n - \frac{\epsilon}{2}\frac{\partial V}{\partial q}(q_n)$
$q_{n+1} \leftarrow q_n + \epsilon \frac{\partial K}{\partial p}(p_{n+\frac{1}{2}})$
$p_{n+1} \leftarrow p_{n+\frac{1}{2}} - \frac{\epsilon}{2}\frac{\partial V}{\partial q}(q_{n+1})$
end for.

$$\frac{\mathrm{d}q}{\mathrm{d}t} = +\frac{\partial H}{\partial p} = \frac{\partial K}{\partial p}$$
$$\frac{\mathrm{d}p}{\mathrm{d}t} = -\frac{\partial H}{\partial q} = -\frac{\partial K}{\partial q} - \frac{\partial V}{\partial q}.$$



Example: Sample a bivariate normal

$$K(q,p) = -\log \pi(p|q) = -\log N(p|0,1))$$
$$V(q) = \frac{1}{2} q^T \Sigma^{-1} q$$

$$\frac{\partial V}{\partial q} = \Sigma^{-1} q \quad \frac{\partial K}{\partial p} = p$$

$q_0 \leftarrow q, p_0 \leftarrow p$
for $1 \leq n \leq \lfloor T/\epsilon \rfloor$ **do**
$p_{n+\frac{1}{2}} \leftarrow p_n - \frac{\epsilon}{2}\Sigma^{-1}(q_n)$
$q_{n+1} \leftarrow q_n + \epsilon\, p_{n+\frac{1}{2}}$
$p_{n+1} \leftarrow p_{n+\frac{1}{2}} - \frac{\epsilon}{2}\Sigma^{-1}(q_{n+1})$
end for.

Hamiltonian Monte Carlo – The Algorithm

The Algorithm:

1. set $t = 0$
2. generate an initial position state $x^{(0)} \sim \pi$
3. repeat until $t = M$
   I. set $t = t + 1$
   II. sample a new initial momentum variable from the canonical momentum distribution $p_0 \sim P(p)$
   III. set $x_0 = x^{(t-1)}$
   IV. run Leap **Frog** algorithm starting at $[x_0, p_0]$ for L steps and step size δ to obtain proposed states x^* and p^*
   V. calculate the Metropolis acceptance probability $\alpha = \min\left(1, \exp\left(-U(x^*) + U(x_0) - K(p^*) + K(p_0)\right)\right)$
   VI. draw a random number u from $Unif(0,1)$
      i. if $u \leq \alpha$ accept the proposed state position x^* and set the next state in the Markov chain $x^{(t)} = x^*$
      ii. else set $x^{(t)} = x^{(t-1)}$

Hamiltonian Monte Carlo – The Movie

https://chi-feng.github.io/mcmc-demo/app.html#HamiltonianMC,banana

References

- https://stats.stackexchange.com/questions/261796/relationship-between-hessian-matrix-and-covariance-matrix
- https://arxiv.org/abs/1206.1901
- https://theclevermachine.wordpress.com/2012/11/18/mcmc-hamiltonian-monte-carlo-a-k-a-hybrid-monte-carlo/
- http://khalibartan.github.io/MCMC-Hamiltonian-Monte-Carlo-and-No-U-Turn-Sampler/
- https://en.wikipedia.org/wiki/Lagrangian_mechanics
- https://www1.maths.leeds.ac.uk/~voss/projects/2011-RJMCMC/AiJialin.pdf
- https://stats.stackexchange.com/questions/271385/mcmc-metropolis-hasting-formal-derivation-of-detailed-balance
- https://www.math.wustl.edu/~sawyer/hmhandouts/MetropHastingsEtc.pdf
- http://dept.stat.lsa.umich.edu/~kshedden/Courses/Stat606/Notes/mcmc.pdf
- https://arxiv.org/pdf/1701.02434.pdf
- https://en.wikipedia.org/wiki/Examples_of_Markov_chains
- https://en.wikipedia.org/wiki/Markov_chain#Stationary_distribution
- https://cims.nyu.edu/~holmes/teaching/asa17/handout-Lecture3_2017.pdf
- https://en.wikipedia.org/wiki/Multivariate_normal_distribution
- https://en.wikipedia.org/wiki/Riemann%E2%80%93Stieltjes_integral
- https://en.wikipedia.org/wiki/Measure_(mathematics)
- https://engineering.purdue.edu/~ipollak/ece302/SPRING12/notes/26_Limit_Theorems_1_WLLN_packed.pdf
- https://en.wikipedia.org/wiki/Law_of_large_numbers
- http://www.ics.uci.edu/~johnsong/papers/Chib%20and%20Greenberg%20-%20Understanding%20the%20Metropolis-Hastings%20Algorithm.pdf

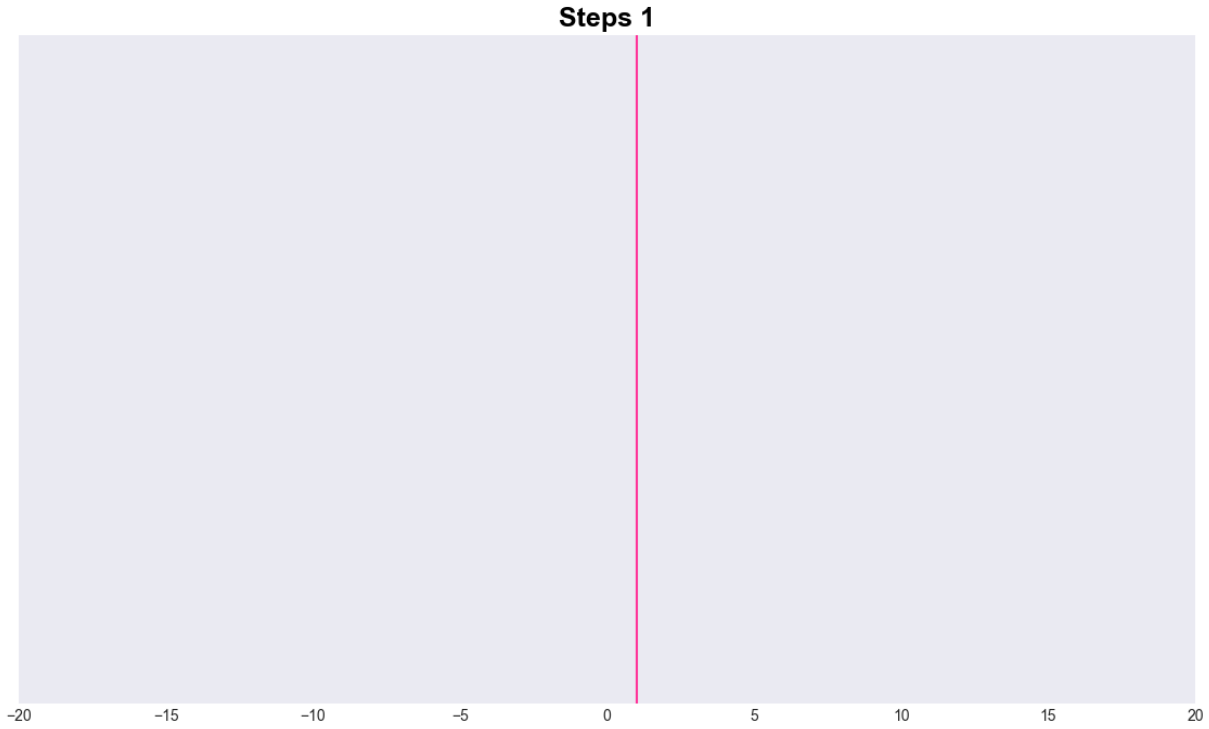
Histograms via Random Walks

MCMC - Metropolis Hastings

```
for i in range(n_samples):

    # propose next step
    x_prop = norm(x_current,std).rvs()

    # build acceptance probabilities
    p_x = (1/area)*f(x_current)
    q_x_prop_given_x = norm(x_current,std).pdf(x_prop)
    p_x_prop = (1/area)*f(x_prop)
    q_x_given_x_prop = norm(x_prop,std).pdf(x_current)
    A_x_x_prop = min(1.0,(p_x_prop*q_x_given_x_prop)/(p_x*q_x_prop_given_x))

    # accept or reject
    u = np.random.rand()
    if u < A_x_x_prop:

        # if proposal accepted then move to the next spot
        samples.append(x_prop)
        x_current = x_prop

    else:

        # if proposal rejected sample given point again
        samples.append(x_current)
```

Distribution Building via Random Walks

Random Walker – start with a sampling algorithm and compute a density f of the distribution of observations

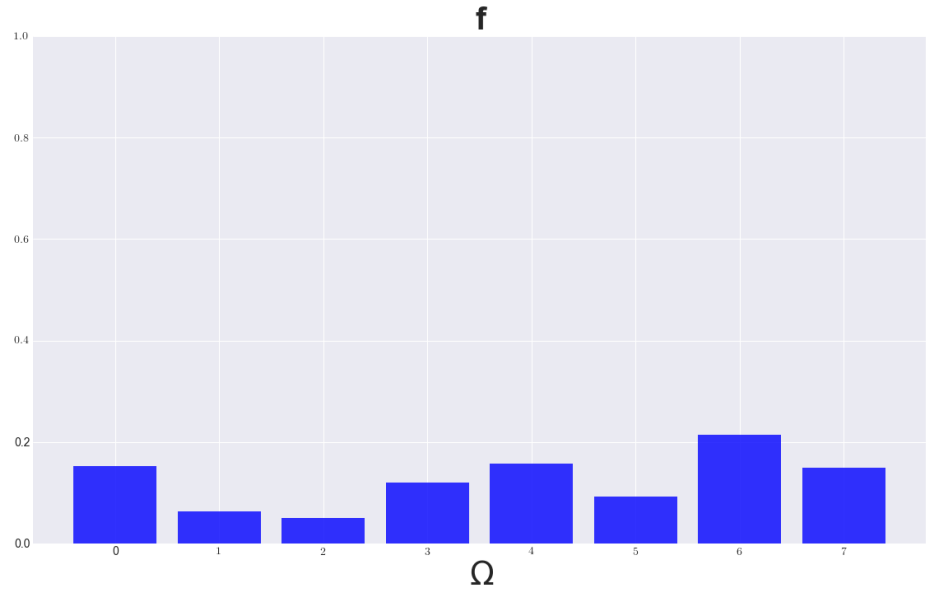


Our Goal: Reverse this – Create a random walker in such a way as the resulting density matches the prescribed function up to multiplication by a scalar

Distribution Building via Random Walks

Example: Intuit a Markov method of sampling from histogram

Consider a normalized histogram f defined over $\Omega = \{0,1, \dots, 7\}$



Let S_N denote our current position in $\{\mathrm{x_i}\} \coloneqq \{0,1, \dots, 7\}$. Need a procedure for choosing S_{N+1}.

Proposal:

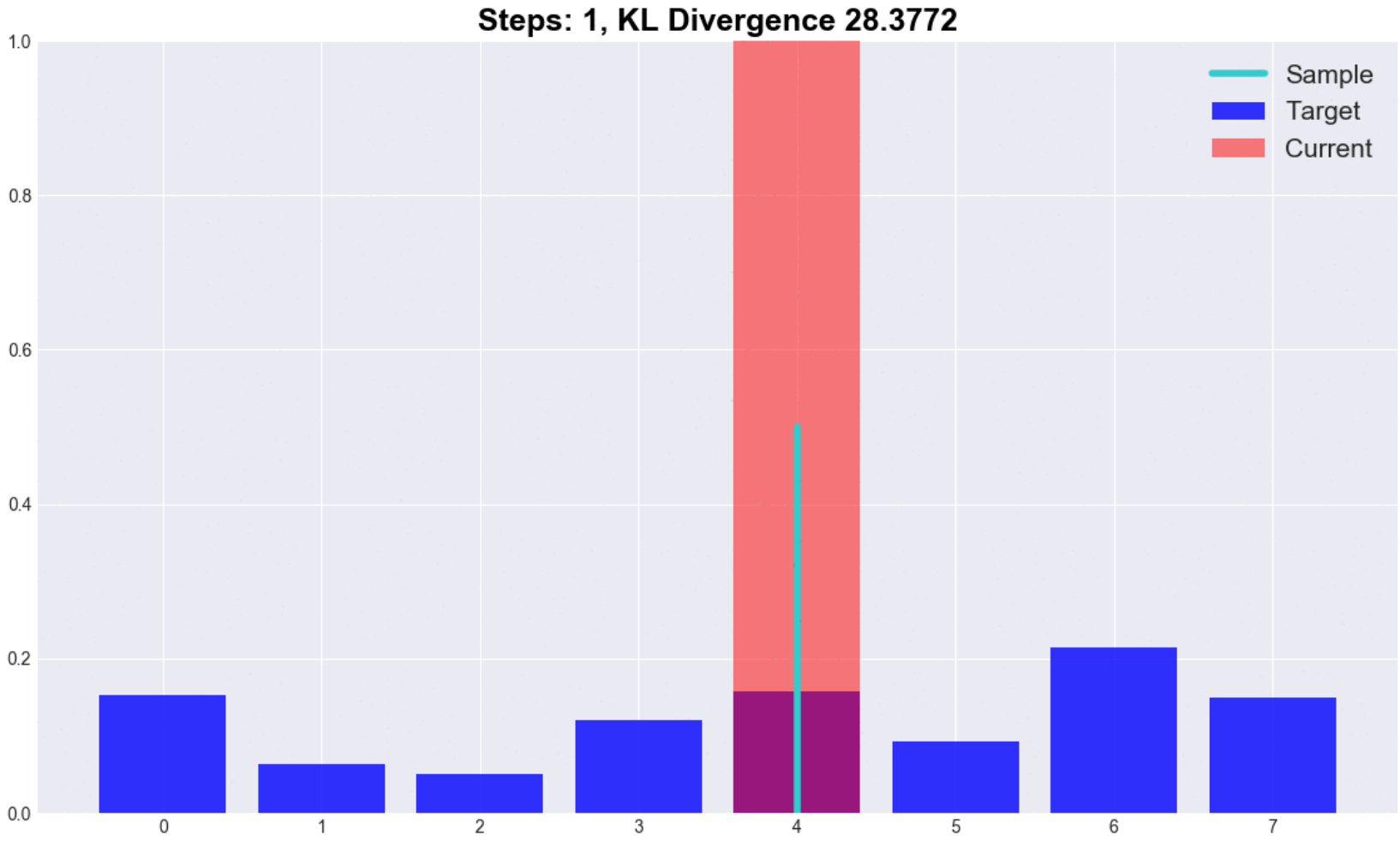
$$prop_i \sim U(x_1, \mathrm{x}_2, \dots, x_{i-1}, x_{i+1}, \dots, x_N)$$

Acceptance:

$$P\left(accept\ x_i \middle| x_j\right) = \min\left(\frac{f(x_i)}{f(x_j)}, 1\right)$$

Need to define a random walk procedure in such a way that the histogram of counts of steps on elements of Ω matches f

Distribution Building via Random Walks



The Typical Set: Example

MV Gaussian $\mu = 0, \Sigma = I_n$