Hamiltonian Monte Carlo

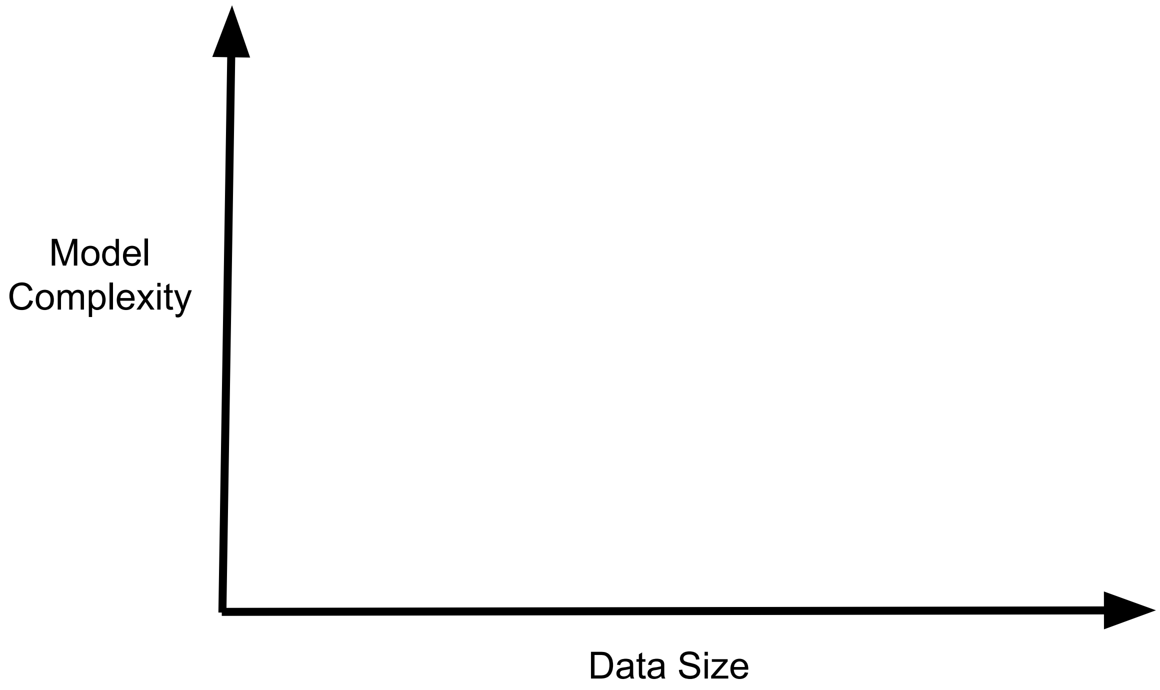
https://github.com/adityanemali/journal_club_dzne

Motivation

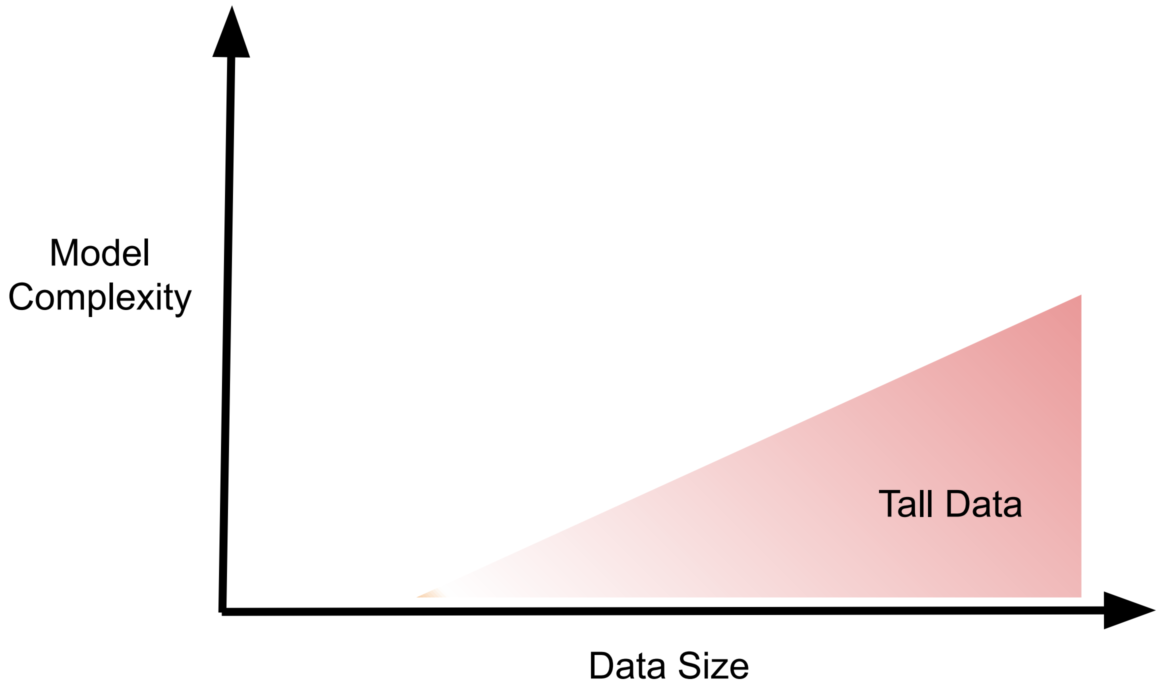
In order to build a complete analysis we need to compliment our data with a *model* of its structure

Data Size

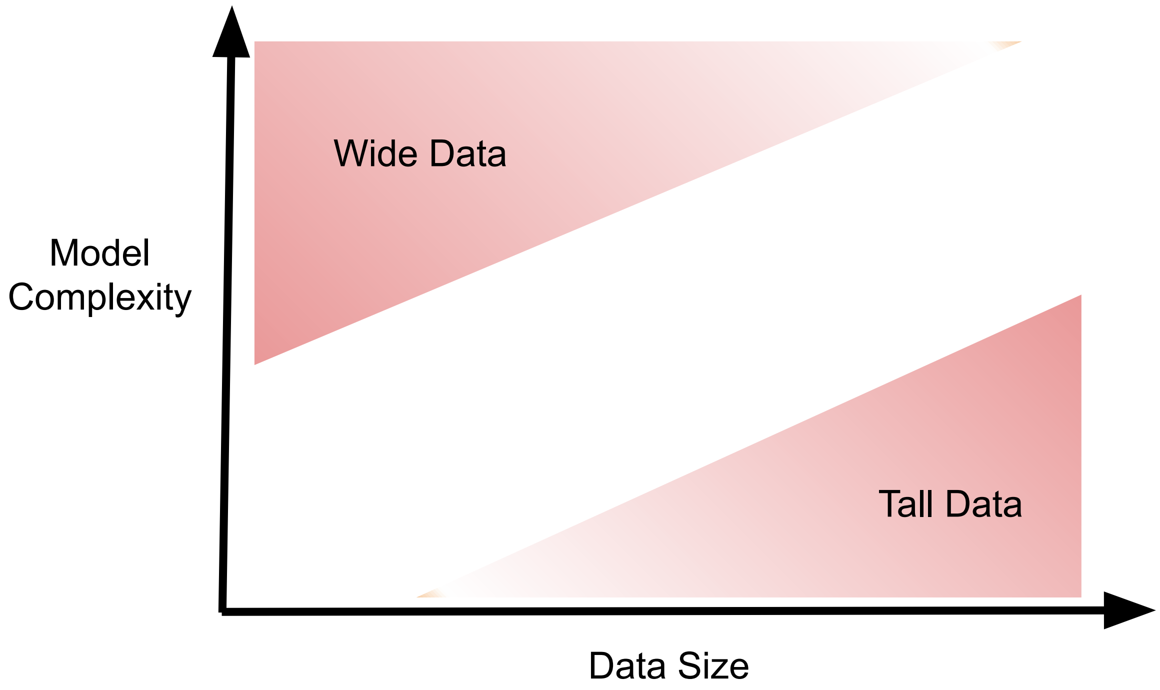
In order to build a complete analysis we need to compliment our data with a *model* of its structure



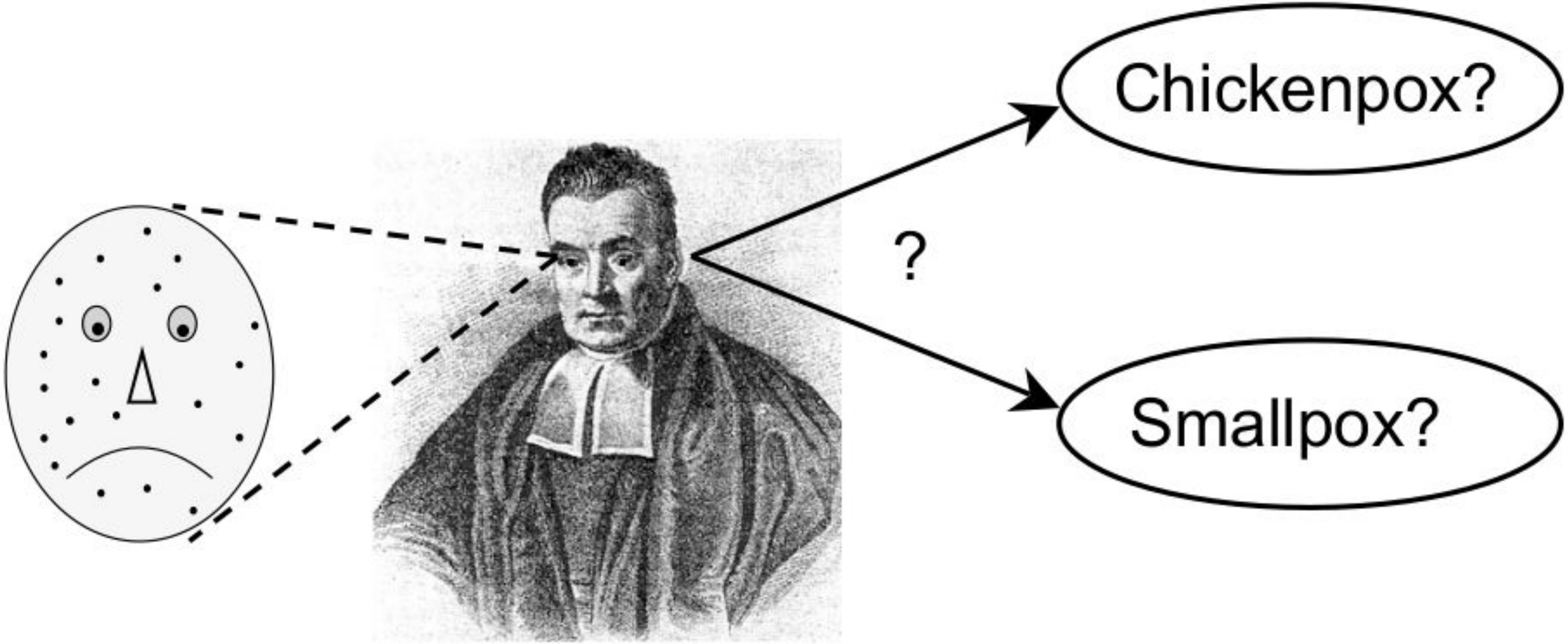
In order to build a complete analysis we need to compliment our data with a *model* of its structure



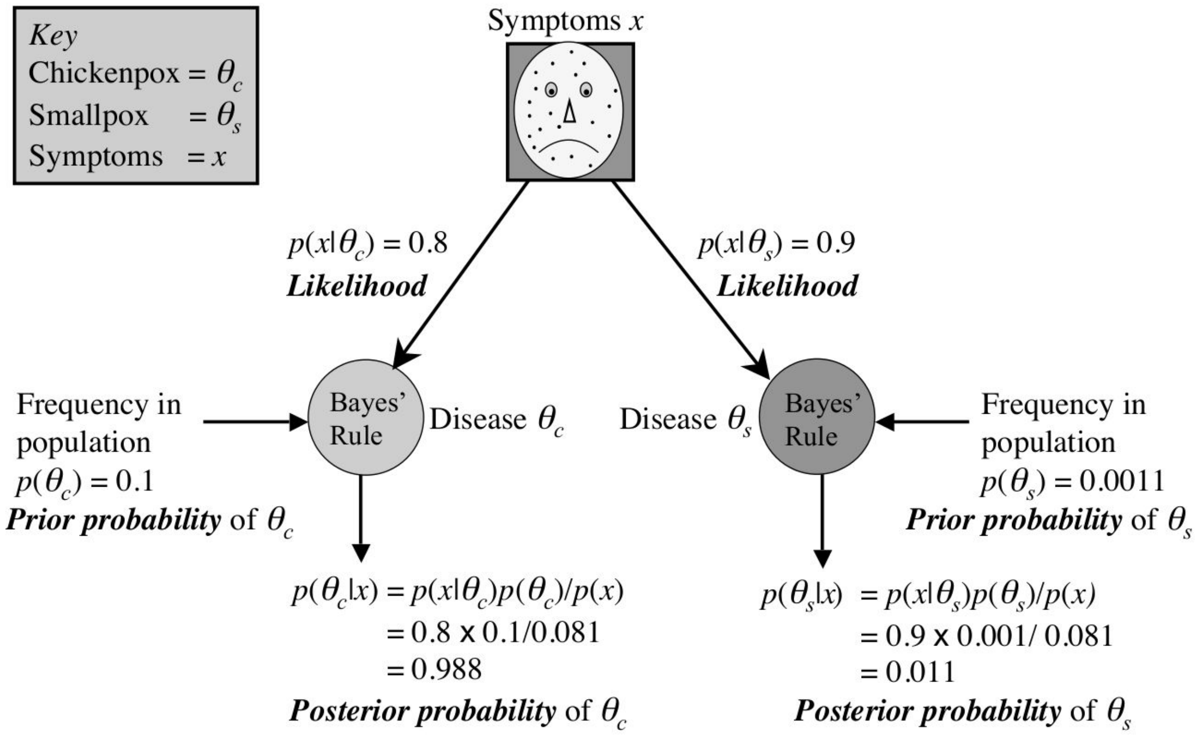
In order to build a complete analysis we need to compliment our data with a *model* of its structure



Bayesian inference is a powerful tool for asking statistical questions



Chickenpox?
?
Smallpox?


Key
Chickenpox = θ_c
Smallpox = θ_s
Symptoms = x
Symptoms x
$p(x|\theta_c) = 0.8$
Likelihood
$p(x|\theta_s) = 0.9$
Likelihood
Frequency in population
$p(\theta_c) = 0.1$
Prior probability of θ_c
Bayes' Rule
Disease θ_c
Disease θ_s
Bayes' Rule
Frequency in population
$p(\theta_s) = 0.0011$
Prior probability of θ_s
$p(\theta_c|x) = p(x|\theta_c)p(\theta_c)/p(x)$
$= 0.8 \times 0.1/0.081$
$= 0.988$
Posterior probability of θ_c
$p(\theta_s|x) = p(x|\theta_s)p(\theta_s)/p(x)$
$= 0.9 \times 0.001/ 0.081$
$= 0.011$
Posterior probability of θ_s

Bayesian Inference

Posterior Distribution ∝ Likelihood * Prior

$$\pi(q|D) \propto \pi(D|q) * \pi(q)$$

But what makes a good statistical question?

Computing Expectations

$$\mathbb{E}_{\pi}[f] = \int_{Q} \mathrm{d}q \, \pi(q) \, f(q) \, .$$

Parsimonious Expectation Computation

Prob. Computational inefficiency - waste resources evaluating the target density and relevant functions that have negligible contribution to the desired expectation.

Sol. Avoid those regions - we first need to identify how the target density and the target function contribute to the overall expectation.

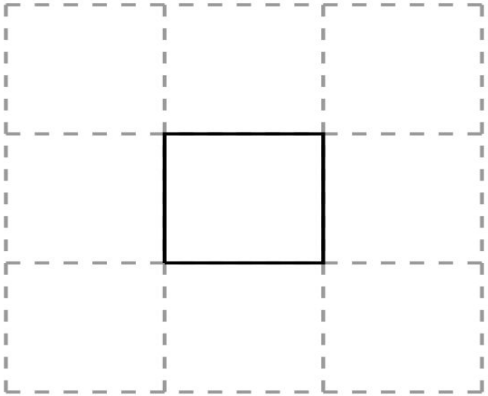
Parsimonious Expectation Computation

We consider regions where the target density functions take larger values.

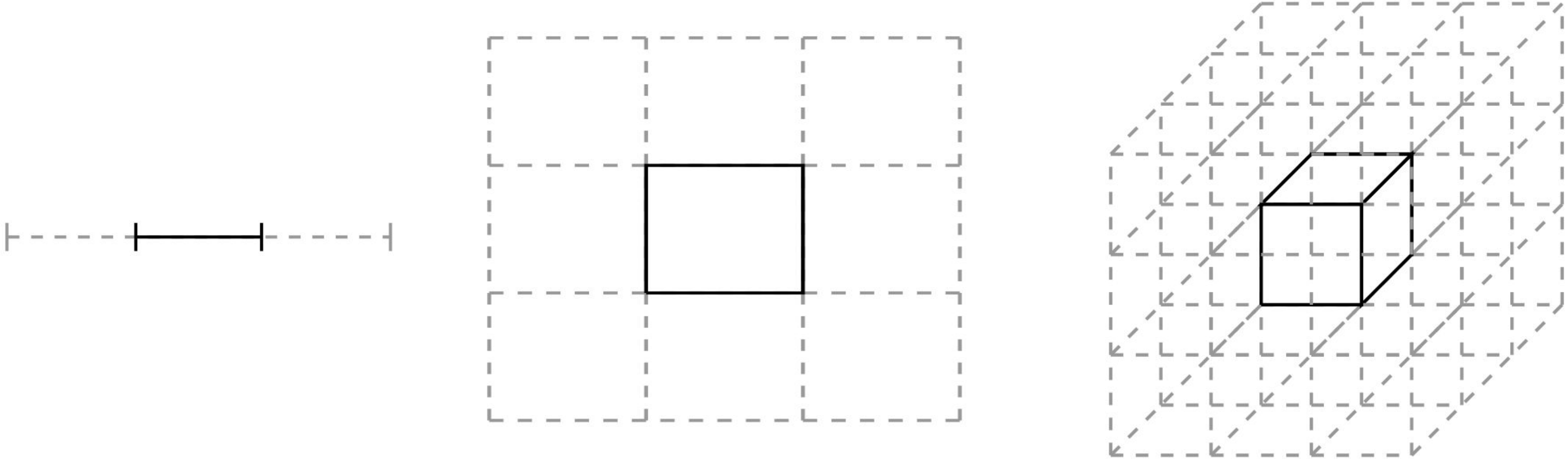
$$\pi(q)\, f(q)$$

Volume, however, starts to behave strangely as the dimensions of our parameter space increases.

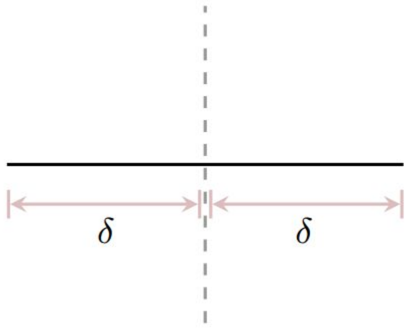
Volume, however, starts to behave strangely as the dimensions of our parameter space increases.



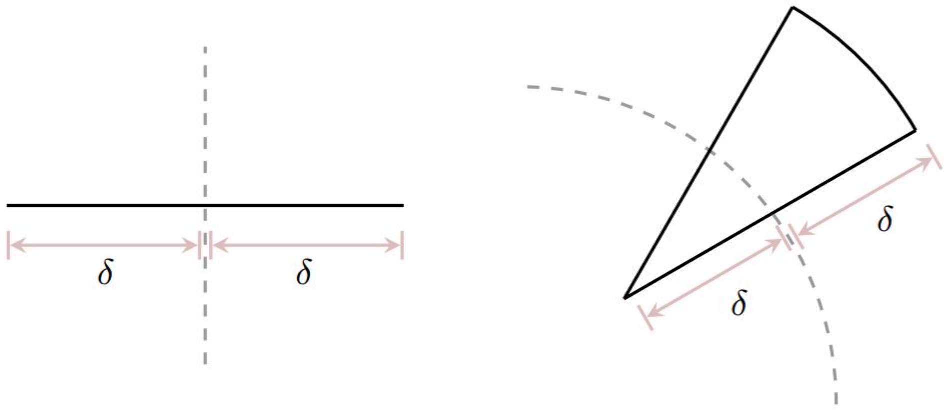
Volume, however, starts to behave strangely as the dimensions of our parameter space increases.



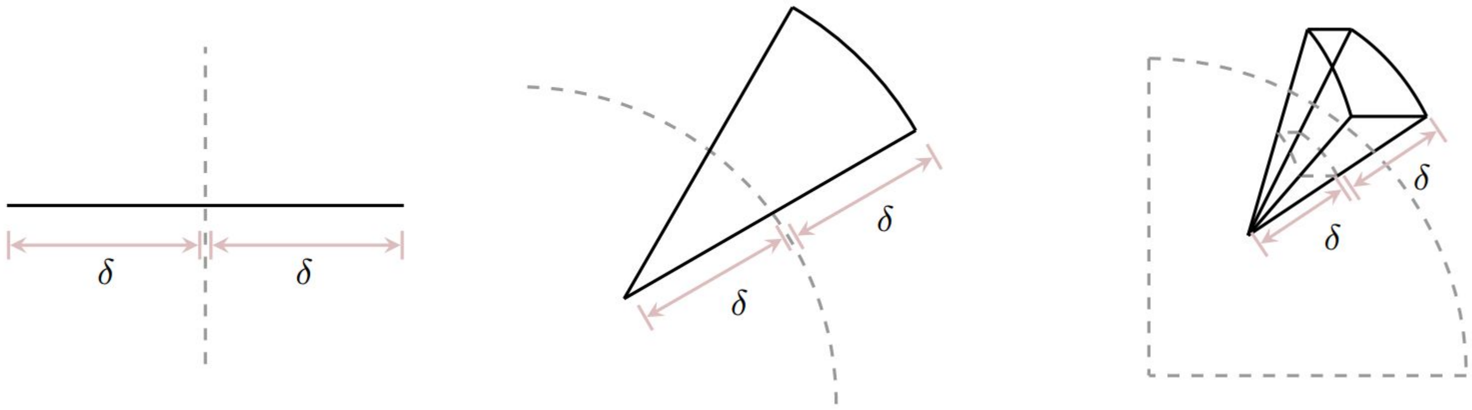
Volume, however, starts to behave strangely as the dimensions of our parameter space increases.



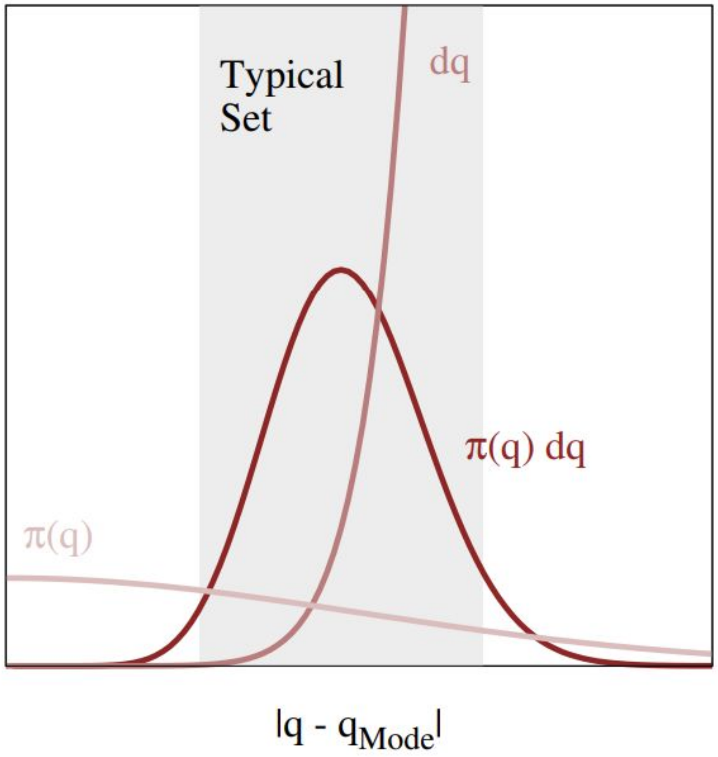
Volume, however, starts to behave strangely as the dimensions of our parameter space increases.



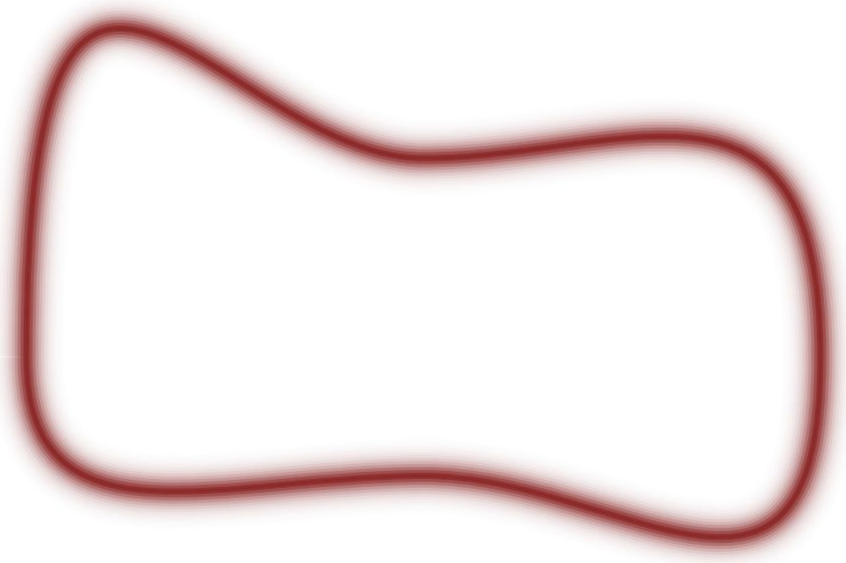
Volume, however, starts to behave strangely as the dimensions of our parameter space increases.



If relevant neighbours are determined by probability density then we should focus computations near mode



This concentration of measure into a narrow typical set frustrates the accurate estimation of integrals



Markov Chains

$$P(X_{n+1} = X | X_n, X_{n-1}, ..., X_1, X_0) = P(X_{n+1} = X | X_n)$$

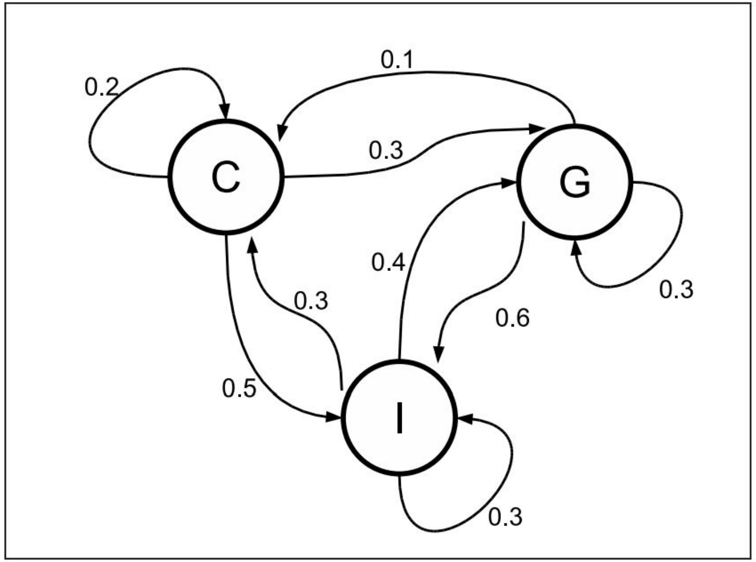


Figure 4: Chinese, Greek or Italian food for lunch?

Markov Chains

in the next days and got the following data from Mrs. Bond:

- probability vector (also initial state vector) $\pi_0 = \begin{pmatrix} \pi_C \\ \pi_G \\ \pi_I \end{pmatrix}_{3x1} = \begin{pmatrix} 1.0 \\ 0.0 \\ 0.0 \end{pmatrix}$
- transition probability matrix (selected from given picture) $P = \begin{pmatrix} 0.2 & 0.3 & 0.5 \\ 0.1 & 0.3 & 0.3 \\ 0.3 & 0.4 & 0.3 \end{pmatrix}_{3x3}$

- solution:

$\pi_{k+1} = \pi_k \cdot P$

auxiliary calculation: $\pi_1 = \pi_0 \cdot P$ $\quad\downarrow \begin{pmatrix} 0.2 & 0.3 & 0.5 \\ 0.1 & 0.3 & 0.3 \\ 0.3 & 0.4 & 0.3 \end{pmatrix}_{\underline{3}x3}$

$\rightarrow$

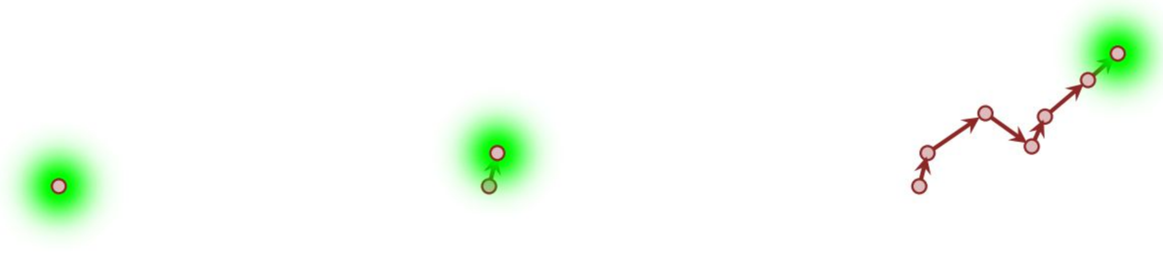
$\pi_0{}^T = (1.0\ 0.0\ 0.0)_{1x\underline{3}}$ $\qquad \pi_1{}^T = (0.2\ 0.3\ 0.5)$

$\pi_1{}^T = (0.2000\ 0.3000\ 0.5000)$
$\pi_2{}^T = (0.2200\ 0.3500\ 0.4300)$
$\pi_3{}^T = (0.2080\ 0.3430\ 0.4490)$
$\pi_4{}^T = (0.2106\ 0.3449\ 0.4445)$
$\pi_5{}^T = (0.2100\ 0.3445\ 0.4456)$
$\pi_6{}^T = (0.2101\ 0.3446\ 0.4453)$
$\pi_7{}^T = (0.2101\ 0.3445\ 0.4454)$
$\pi_8{}^T = (0.2101\ 0.3445\ 0.4454)$

Estimating Expectations with Markov Chains

$$\pi(q) = \int_{\mathcal{Q}} \mathrm{d}q' \, \pi(q') \, \mathbb{T}(q \mid q')$$

Estimating Expectations with Markov Chains



(a) (b) (c)

FIG 5. *(a) A Markov chain is a sequence of points in parameter space generated by a Markov transition density (green) that defines the probability of a new point given the current point. (b) Sampling from that distribution yields a new state in the Markov chain and a new distribution from which to sample. (c) Repeating this process generates a Markov chain that meanders through parameter space.*

Estimating Expectations with Markov Chains

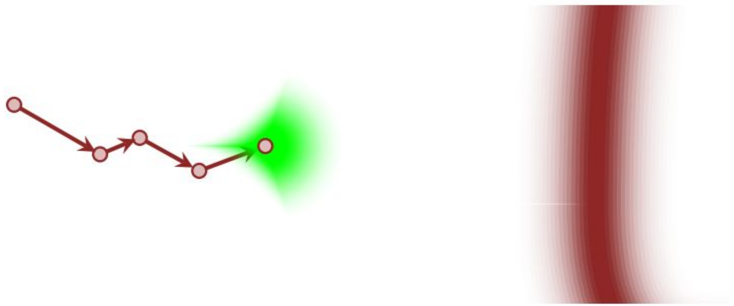


FIG 6. *When a Markov transition (green) preserves the target distribution, it concentrates towards the typical set (red), no matter where it is applied. Consequently, the resulting Markov chain will drift into and then across the typical set regardless of its initial state, providing a powerful quantification of the typical set from which we can derive accurate expectation estimators.*

Here the posterior is represented with a set of samples from which expectations can be efficiently computed

$$\hat{f}_N = \frac{1}{N} \sum_{n=0}^{N} f(q_n)$$

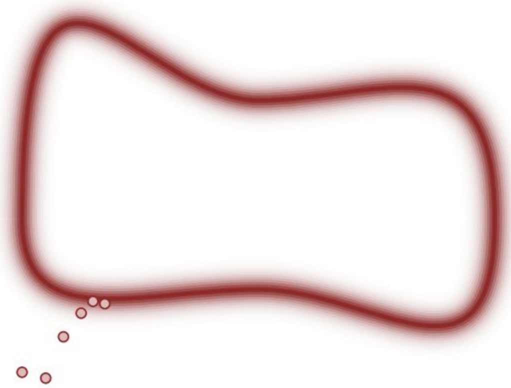
Here the posterior is represented with a set of samples from which expectations can be efficiently computed

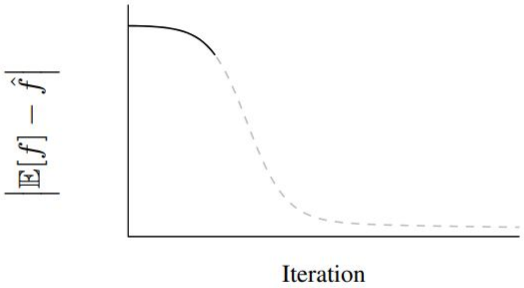
$$\hat{f}_N = \frac{1}{N} \sum_{n=0}^{N} f(q_n)$$

$$\lim_{N \to \infty} \hat{f}_N = \mathbb{E}_\pi[f]$$

Behaviour of Markov Chains

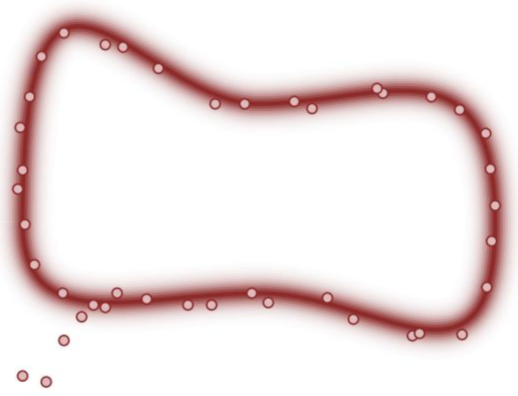
First Phase

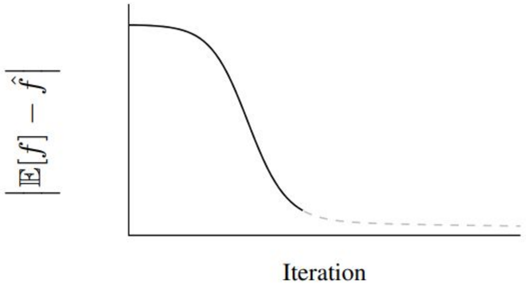




Behaviour of Markov Chains

Second Phase





Behaviour of Markov Chains

Third Phase

$$\hat{f}_N^{\text{MCMC}} \sim \mathcal{N}(\mathbb{E}_\pi[f], \text{MCMC-SE}) \,,$$

where the *Markov Chain Monte Carlo Standard Error* is given by

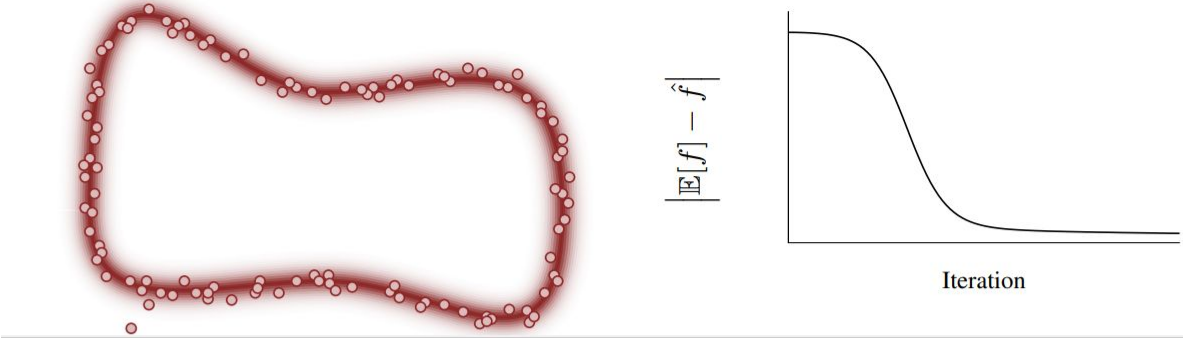
$$\text{MCMC-SE} \equiv \sqrt{\frac{\text{Var}_\pi[f]}{\text{ESS}}}.$$

The *effective sample size* is defined as

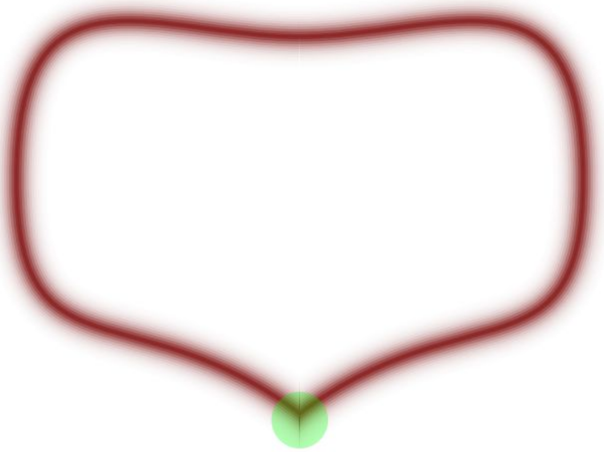
$$\text{ESS} = \frac{N}{1 + 2\sum_{l=1}^{\infty} \rho_l},$$

Behaviour of Markov Chains

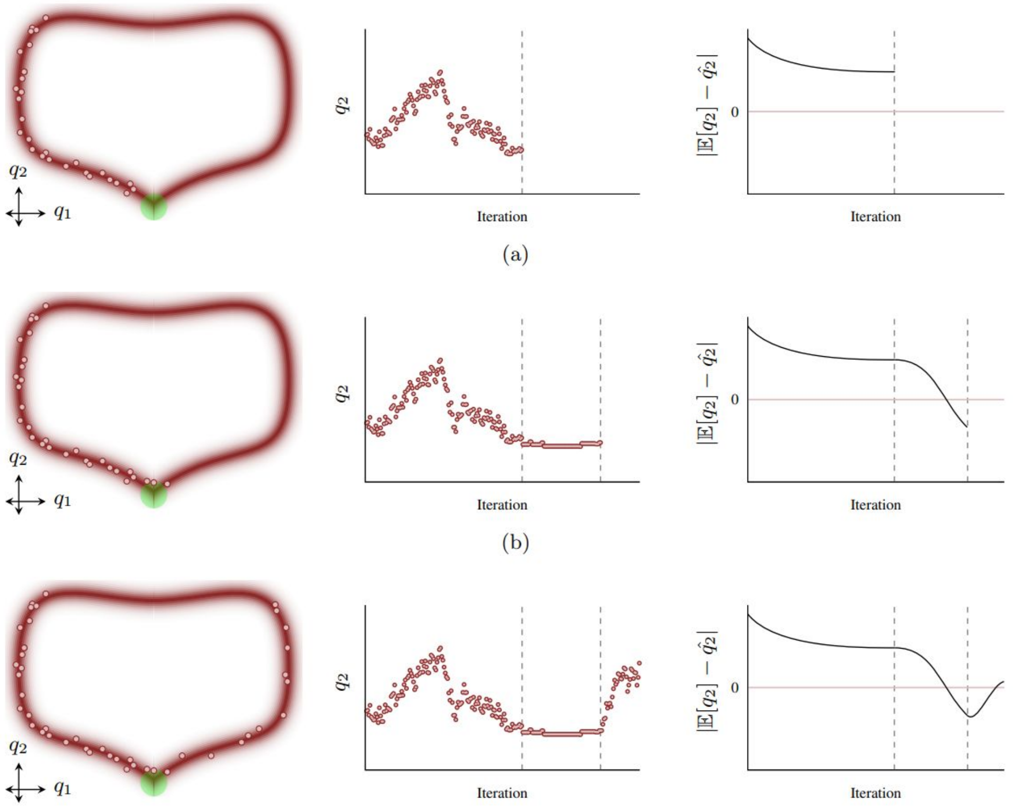
Third Phase



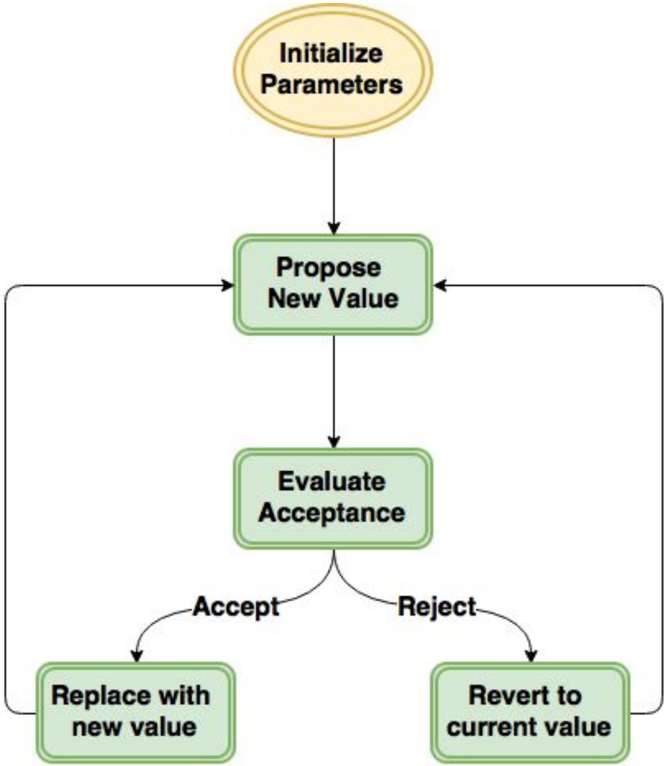
Posterior geometries, spoil the ideal conditions



Posterior geometries, spoil the ideal conditions



Metropolis Hasting Algorithm



Metropolis Hasting Algorithm

Two Phases : Proposal, Correction

$$a(q' \mid q) = \min\left(1, \frac{\mathbb{Q}(q \mid q')\, \pi(q')}{\mathbb{Q}(q' \mid q)\, \pi(q)}\right)$$

How do we choose the Proposal - Random Walk Metropolis

$$\mathbb{Q}(q' \mid q) = \mathcal{N}(q' \mid q, \Sigma),$$

Metropolis Hasting Algorithm

Random Walk Metropolis

$$a(q' \mid q) = \min\left(1, \frac{\pi(q')}{\pi(q)}\right)$$

Questions?