Date: Thursday, February 7, 2019

* **Aim now:** Estimating posterior distributions using Markov chain Monte Carlo (MCMC).

* We will begin today by looking into Monte Carlo methods and then we'll add Markov chains tomorrow.

First motivating example: RGB images

$\boxed{\text{RGB}} \rightarrow \boxed{\text{Grayscale}} \rightarrow \boxed{P_{ij}}$ } P matrix

P_{ij} = Pixel intensity (Between 0 and 1) of coordinate (i,j)

QUESTION: Can we sample new images from the matrix (distribution) P?

ANSWER: Yes, it's easy when we know P.
↳ Monte Carlo methods & Rejection Sampling.

QUESTION: What if we didn't know P exactly but had some guess as to what it is?

Example: Perhaps P was large in memory, so we compress it to some lower dimensional space (say S via PCA,
Spectral Clustering,
Fourier Transformation,
Neural Networks,
wavelet, etc.)
Then, we get our best guess of P with $Q = f(S)$.

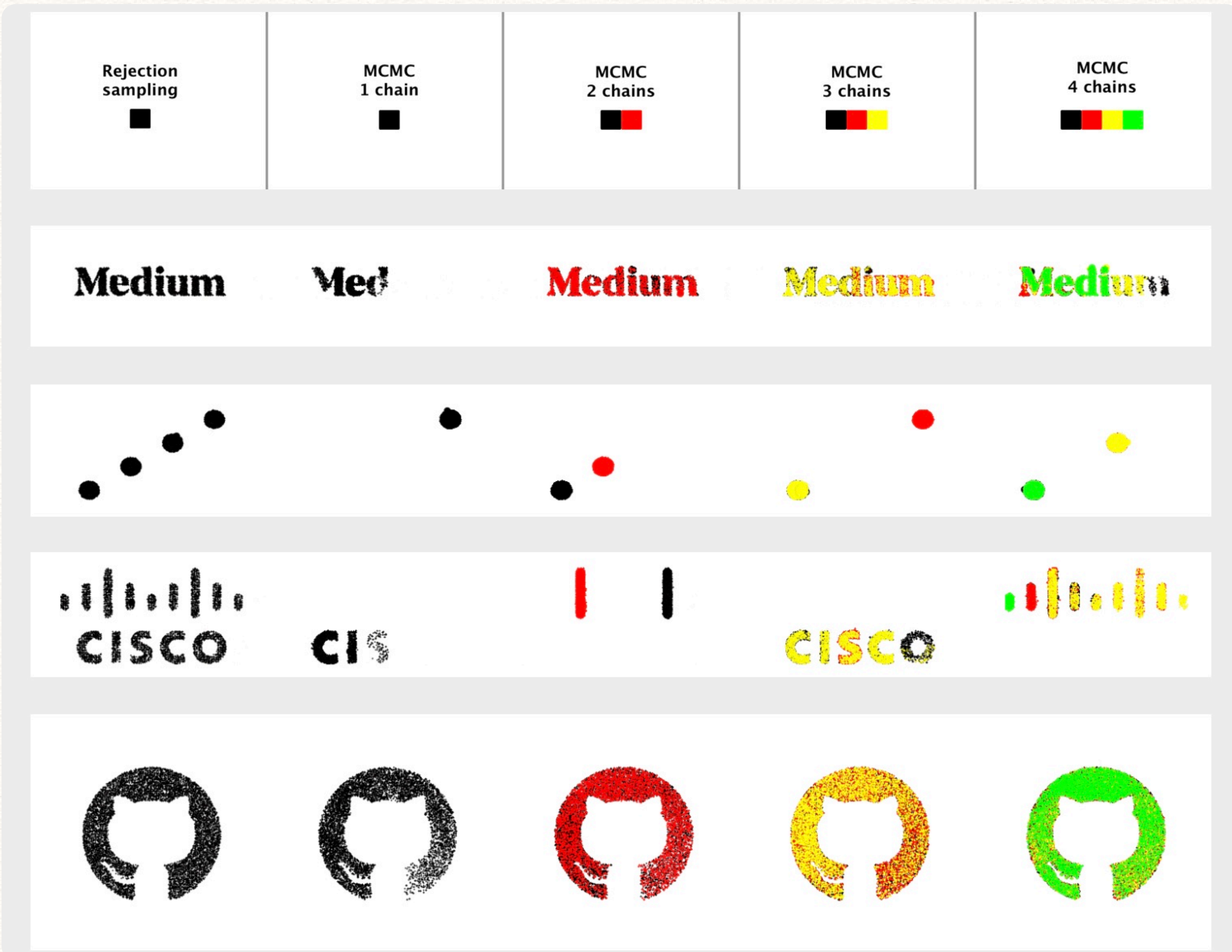
Q: Can we still sample from P using Q?
A: Yes! Using MCMC!

Quick Note: Monte Carlo methods are in general faster than MCMC methods.

The reason is due to the fact they solve easier questions.

MCMC → used for partial info about P.

MC → used when we know P.



[Slides time]

Monte Carlo Simulations

Motivating Example: $f(x) = \frac{1}{1+e^{-x}}$ Sigmoid function

TASK: Simulate values of $f(x)$ from this function.
↳ Common idea: Simulate x's and then get values from $f(x)$.

↳ with just getting the values of a function like $f(x)$, this is easy → draw function

* when $f(x)$ is a density function which weights the likelihood of a random variable, this is not easy.

Example: x is a random variable w/ density $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. Simulate values of X.

↳ If you don't know, this is the density of a $N(0,1)$, you've got to come up with a smart way to sample.

Target distribution: $\mathbb{P}(\theta \mid y)$ [Posterior]

Unnormalized density: $q(\theta \mid y) = \pi(\theta) f(y \mid \theta)$ [numerator of $\mathbb{P}(\theta \mid y)$]

Numerical Integration

⇒ Law of Large numbers (sample means)

* Slide 8/23 pseudo-code:

We know the posterior distribution
$\theta \mid y \sim N(\mu, \sigma^2)$

Remember factorial?

$n! = n \cdot (n-1) \cdot \ldots \cdot 3 \cdot 2 \cdot 1$

This is him now

$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t}\, dt$

Feel old yet?

¡¡¡ Gamma Function !!!

Task: Approximate $\mathbb{E}[\log(|\theta|) \mid y]$

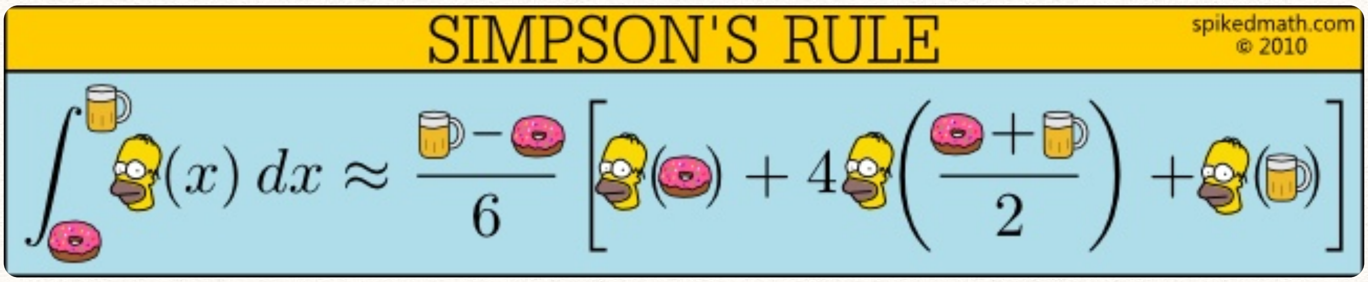
1) Sample S samples of θ from posterior:

   $\theta^{(s)}$ = a draw from $N(\mu, \sigma^2)$ assumes we know how to sample from $N(\mu, \sigma^2)$

2) Plug in w/ sample mean:

   $$\mathbb{E}[\log(|\theta|) \mid y] \approx \frac{1}{S} \sum_{s=1}^{S} \log(|\theta^{(s)}|)$$

Deterministic Methods for Numerical Integration



Rejection Sample

* We know the function $\mathbb{P}(\theta \mid y)$ but don't know how to sample from it.

IDEA: Pick a proposal function $g(\theta)$ that we do know how to simulate from!

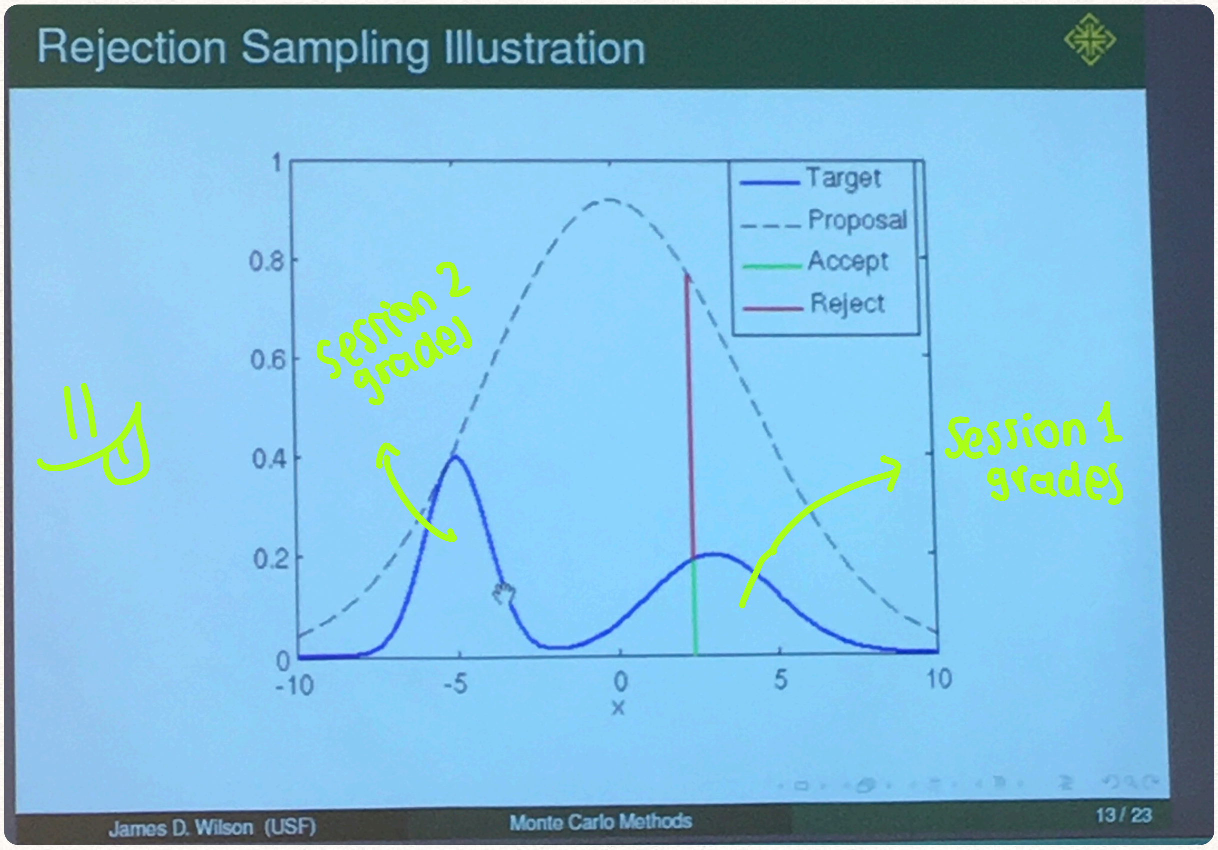
$g(\theta)$ must satisfy:

The function is always "above"

1) $\dfrac{\mathbb{P}(\theta \mid y)}{g(\theta)} \leq M \iff \mathbb{P}(\theta \mid y) \leq M\, g(\theta) \quad \forall \theta$

   importance ratio

2) $g(\theta)$ is integrable (if we know how to simulate from it, this is given).



[Stop point: Slide 16/23]

⇒ Jeff Hamrick : Rejection Sampling
(Wolfram Alpha / YouTube)