Math76 Summer 2020

Introduction to Bayesian Computation

19

Lecture ~~17~~:

MCMC Accuracy

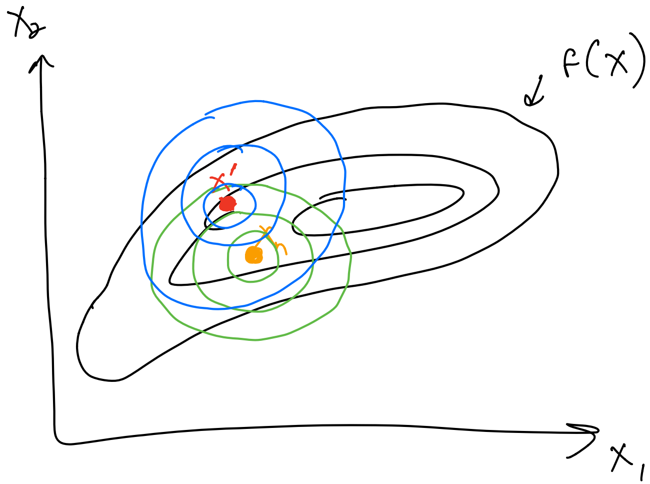
10 August 2020

Simplification with Symmetric Proposals

$$\gamma = \frac{f(x')}{f(x_n)} \frac{q(x_n \mid x')}{q(x' \mid x_n)}$$

$$q(x \mid x') = N(x', \sigma^2)$$

$$q(x \mid x_n) = N(x_n, \sigma^2)$$



Subscripts for component of X

when symmetric

$$q(x_n \mid x') = q(x' \mid x_n)$$

$$\Rightarrow \gamma = \frac{f(x')}{f(x_n)}$$

Monte Carlo Summary

Big Picture:

1. We want to estimate expectations of the form
$$\mathbb{E}[h(X)] = \int_{\Omega_x} h(x) f(x) dx$$

2. In high dimensions or for complicated $h(x)$, Monte Carlo methods can be advantageous
$$\mathbb{E}[h(X)] \approx \frac{1}{N} \sum_{k=1}^{N} h\left(x^{(k)}\right)$$

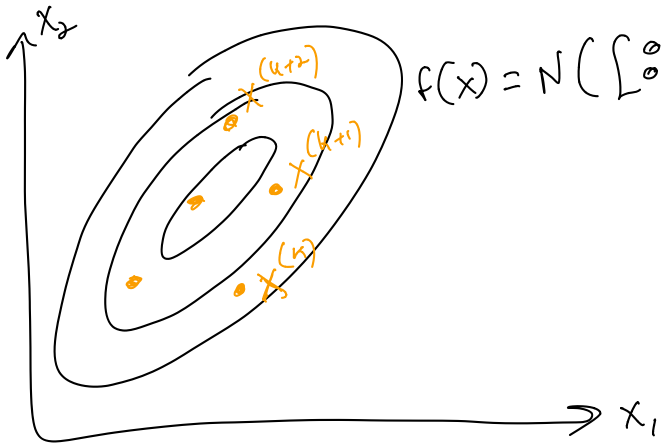
3. Markov chain Monte Carlo is a way of generating samples $x^{(k)}$ using only unnormalized density evaluations.

From "target density" $f(x) \Rightarrow$ samples $x^{(1)}, x^{(2)}, \ldots, x^{(N)}$
of
$f(x)$

Simple Test Problem

X with sample space $\Omega_X = \mathbb{R}^2$

$$X \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \right)$$



$$f(x) = N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \right)$$

$$\mathbb{E}[X_1] = 0 \approx \frac{1}{N} \sum_{k=1}^{N} X_1^{(k)} = \hat{\mu}_N$$

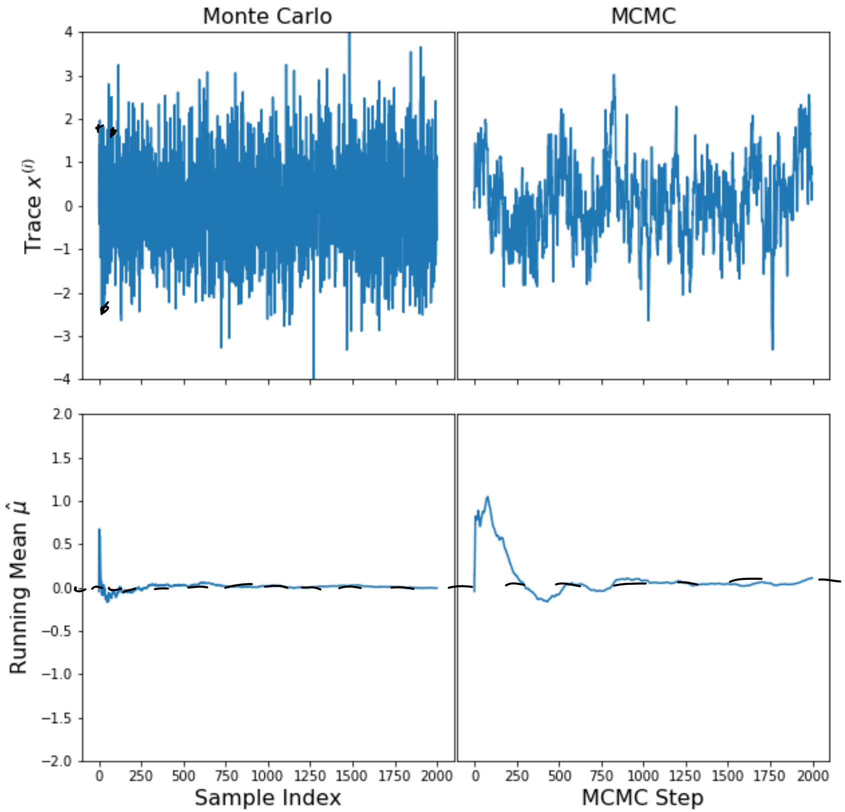
Sample Mean

population mean

"True Expectation"

MC and MCMC on test problem

$N = 2000$

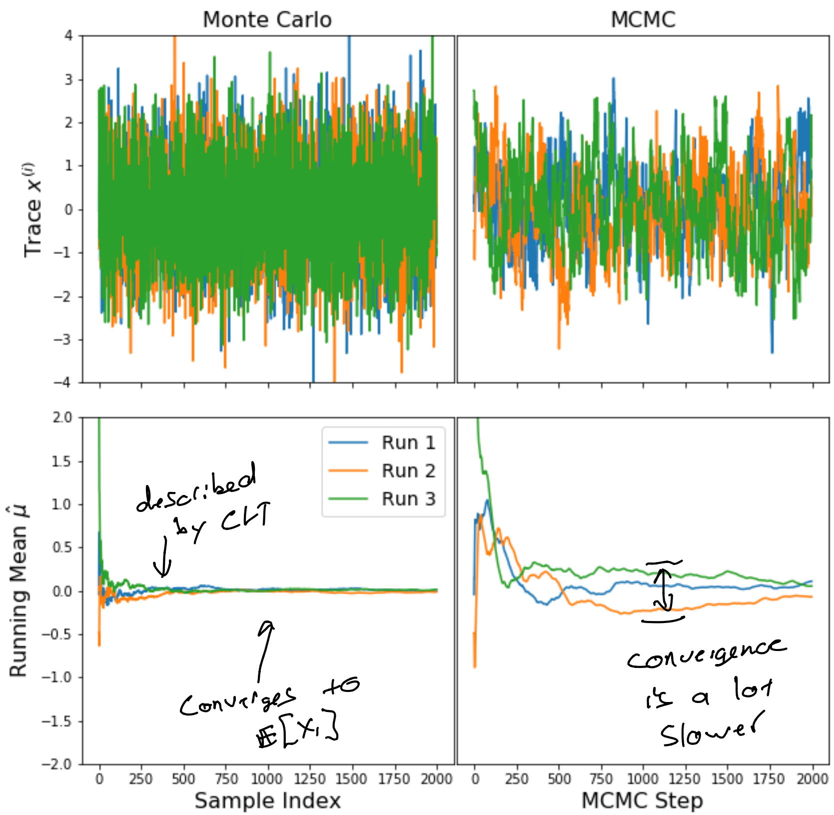


sample mean $\hat{\mu}_N$

$\mathbb{E}[X_1]$

$$\hat{\mu}_N = \frac{1}{N} \sum_{k=1}^{N} x_1^{(k)}$$

MC and MCMC on test problem



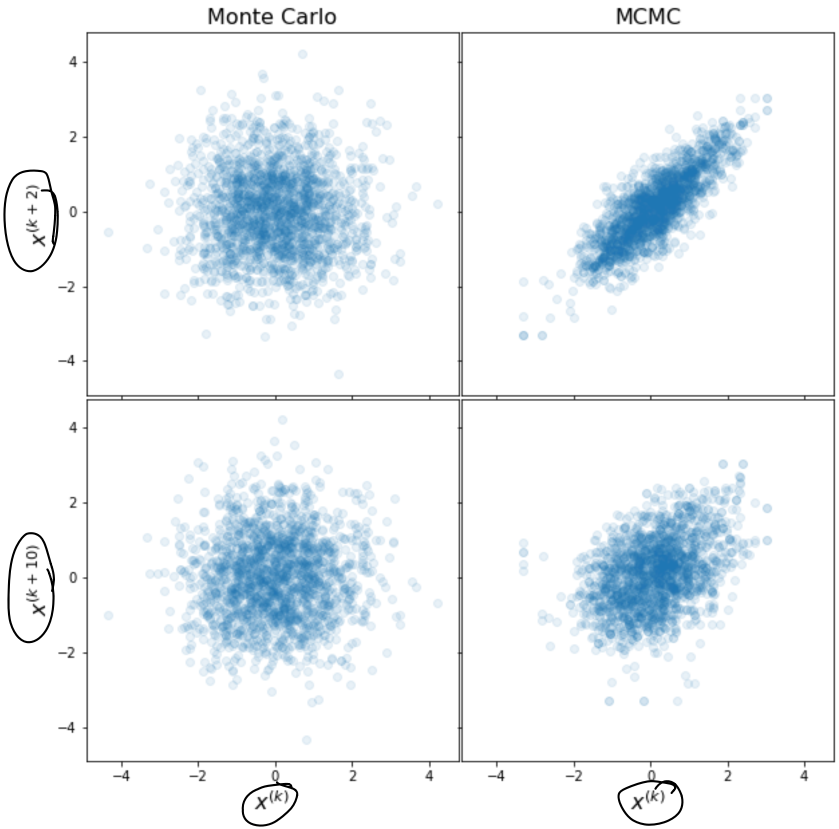
Question:

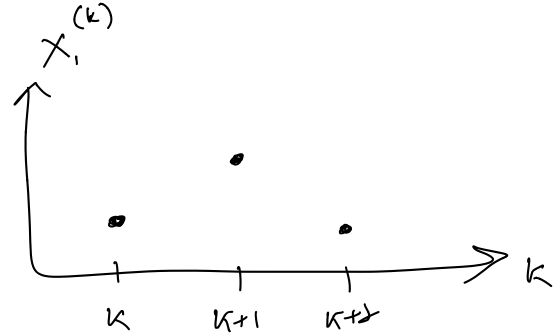
Can we understand MCMC convergence & choose $q(\ )$ to converge as fast as possible.

Recall

From CLT, "convergence" is measured by $\mathrm{Var}[\hat{\mu}_N]$

Chain Correlation





$X_1^{(k)}$ is a RV

$X_1^{(k+d)}$ is a RV

scatterplots comparing $X_1^{(k)}$ & $X_1^{(k+d)}$

Breakout Exercise

Recall the Metropolis-Hastings MCMC algorithm:

1. Choose an initial point z_0.
2. For $i = 1 \ldots N$:
   - Propose a point $z' \sim q(z|z_{i-1})$
   - Compute the acceptance ratio

   $$\gamma = \frac{f(z'|c_{obs})}{f(z_{i-1}|c_{obs})} \frac{q(z_{i-1}|z')}{q(z'|z_{i-1})}$$

   - Set $x_i = x'$ with probability $\alpha = \min\{1, \gamma\}$, else $x_i = x_{i-1}$.

Question:

What two parts of this algorithm cause the chain to be correlated?

↳ why might $x^{(k)}$ & $x^{(d+k)}$ be similar?

Breakout Solution

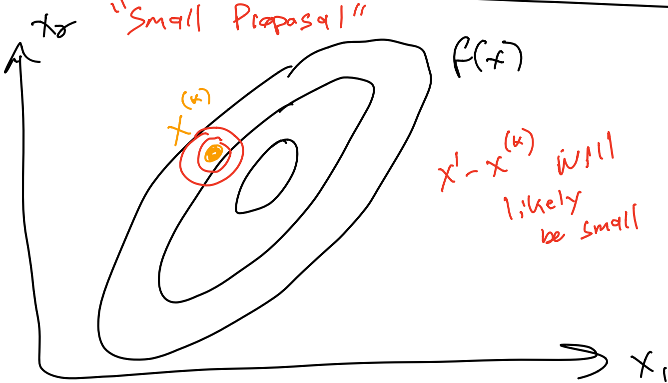
Metropolis-Hastings MCMC algorithm:

1. Choose an initial point z_0.
2. For $i = 1 \dots N$:
   - Propose a point $z' \sim q(z|z_{i-1})$
   - Compute the acceptance ratio

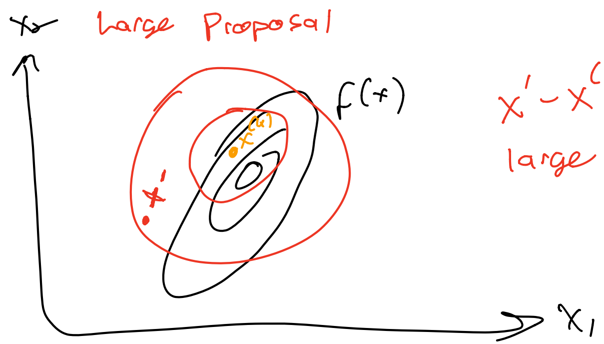
$$\gamma = \frac{f(z'|c_{obs})}{f(z_{i-1}|c_{obs})} \frac{q(z_{i-1}|z')}{q(z'|z_{i-1})}$$

   - Set $x_i = x'$ with probability $\alpha = \min\{1, \gamma\}$, else $x_i = x_{i-1}$.

Small q or very large results in small or zero move



$x' - x^{(k)}$ will likely be small



$x' - x^{(k)}$ is likely large

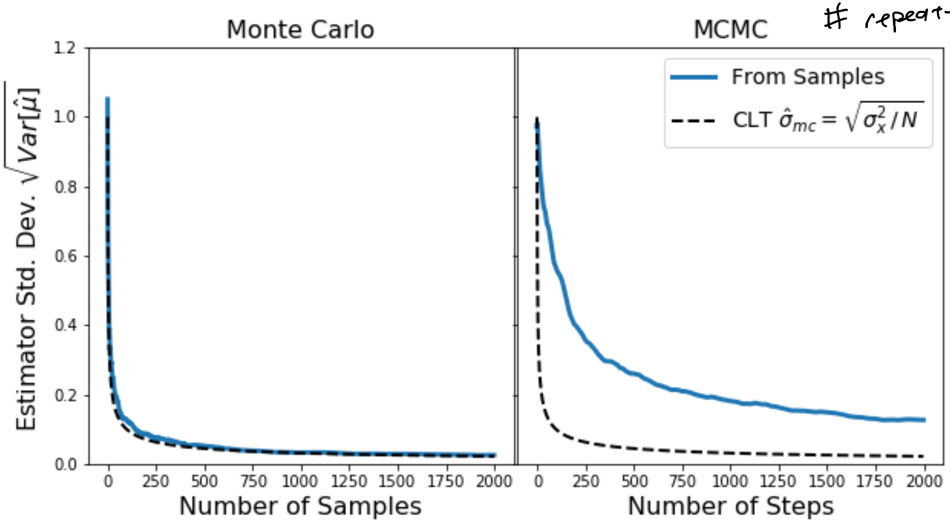
Central Limit Theorem (CLT)

The mean $\hat{\mu}_N$ of independent samples $X^{(k)}$ converges to a Gaussian distribution

$$\frac{1}{N}\sum_{k=1}^{N} X^{(k)} = \hat{\mu}_N \xrightarrow{d} N\left(\mu, \frac{\sigma_X^2}{N}\right)$$

$$\mathrm{Var}[\hat{\mu}_N] \approx \frac{1}{M}\sum_{i=1}^{M} \hat{\mu}_{N,i}$$

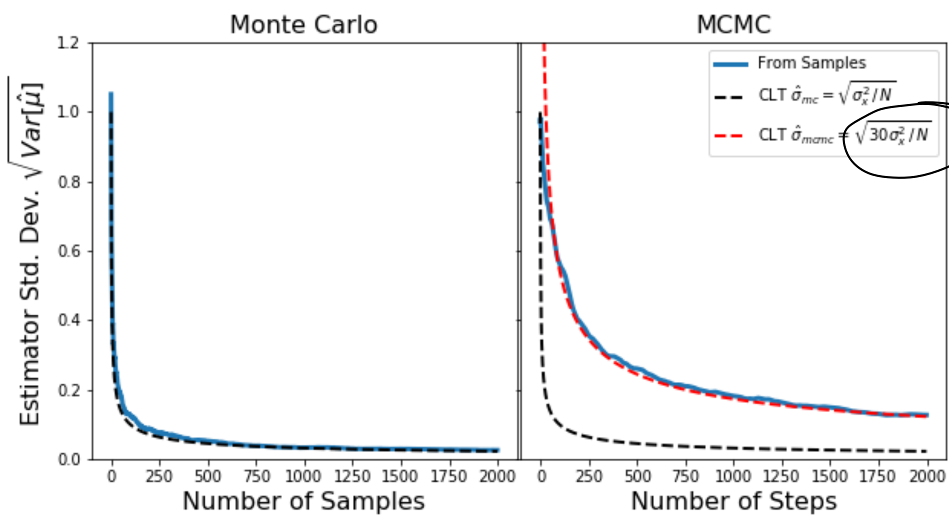
repeated mcmc runs



Central Limit Theorem (CLT)

The mean $\hat{\mu}_N$ of <u>independent</u> samples $X^{(k)}$ converges to a Gaussian distribution

$$\frac{1}{N}\sum_{k=1}^{N} X^{(k)} = \hat{\mu}_N \xrightarrow{d} N\left(\mu, \frac{\sigma_X^2}{N}\right)$$

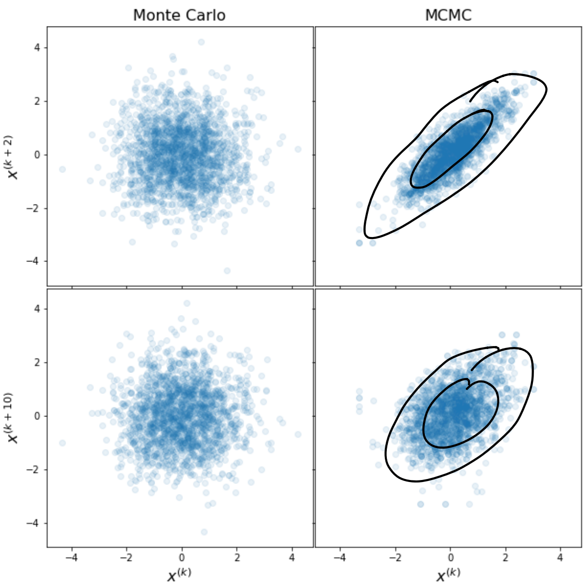


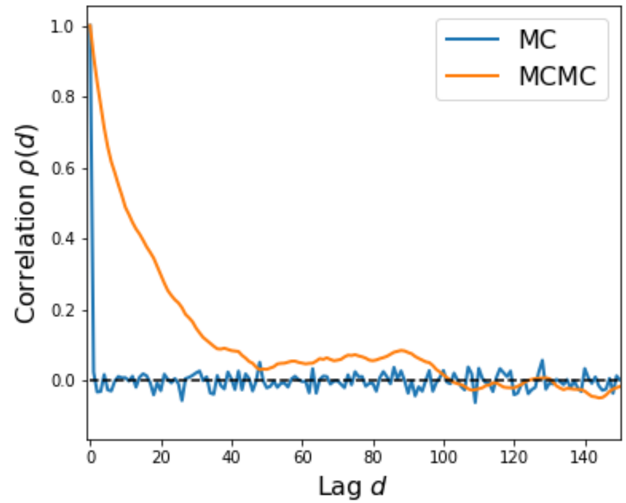
$\frac{30\sigma_x^2}{N}$

30 = Integrated Autocorrelation Time

$\tau. = 30$

Autocorrelation Functions





$$\frac{\mathrm{Cov}\left[x^{(k)}, x^{(k+d)}\right]}{\mathrm{Var}[x] = \mathrm{Cov}\left[x^{(k)}, x^{(k)}\right]} = \rho(d)$$

$\rho(d=0) = 1$

Autocorrelation Function

Integrated Autocorrelation ~~Function~~ Time and CLT

τ = IACT (Integrated AutoCorrelation Time)

$$= \sum_{d=-\infty}^{\infty} \rho(d)$$

$$\approx \sum_{d=-T}^{d=T} \rho(d)$$

correlation function is symmetric $\rho(d) = \rho(-d)$.

$\rho(0) = 1$

$$\Rightarrow \tau \approx 1 + 2\sum_{d=1}^{T} \rho(d)$$

can compute from a single chain

np.correlate( )

CLT w/ MCMC

$$\boxed{Var[\hat{\mu}_N] = \tau \frac{Var[x]}{N}}$$

Minimizing MCMC Error

- Goal: minimize τ (IACT)
  ↳ choosing the proposal q
    ↳ choose variance of Gaussian prop
    ↳ find non-gaussian proposal