TOPIC:

DATE:

FILE UNDER:

PAGE:

(9) Check autocorrelation of your chain (marginally) using acf. Goal: chains should have decreasing acf. Faster decrease indicates faster mixing.

$$acf(1) = \frac{\frac{1}{n}\sum (X_t - \bar{X})(X_{t-1} - \bar{X})}{\frac{1}{n}\sum (X_t - \bar{X})^2}$$

(10) Computing clusters allow easy implementation of multiple chains. So why not?

Simulated Annealing

Suppose we want to minimize a function $h(\theta)$, where θ is a vector of parameters. $\theta \in S$

The basic idea is to simulate a MC with the stationary dist.

MC for simulation

"Annealing"

$$\pi_T(\theta) \propto \exp(-h(\theta)/T)$$

The parameter T is called the "temperature".

Markov Chains for "simulation", annealing (cooling) for minimization.

Let $S^* = \{\theta^* \in S : h(\theta^*) = \min_{\theta} h(\theta)\}$

the set of global minimizers

Define a distribution π^* on S^* so that

$\pi^*(\theta) \propto 1$ for $\theta \in S^*$ and 0 otherwise

TOPIC:

DATE:

FILE UNDER:

PAGE:

~~Thm: As $T \downarrow 0$ we have~~

~~$\pi_T(\theta) \to \pi(\theta) \quad \forall \theta \in S$~~

FACT:

As $T \downarrow 0$, we have $\pi_T \to \pi^*$

Obviously, we'd like to simulate from π^* to get the global minimum, but we can't.

But for a fixed $T > 0$, we can simulate from π_T.

Use Metropolis algorithm with a symmetric proposal density, i.e. ~~$q(\quad|x) = q(x|\quad)$~~

$$q(\theta_n | \theta_{n-1}) = q(\theta_{n-1} | \theta_n)$$

Suppose we are in state n, θ_n,

① Simulate $\theta \sim q(\theta | \theta_n)$

② Simulate $u \sim \text{unif}(0,1)$

③ Set $\theta_{n+1} = \theta$ if $u \le \min\left(\frac{\exp(-h(\theta)/T)}{\exp(-h(\theta_n)/T)}, 1\right)$

$$= \min\left(\exp\left(-\underbrace{(h(\theta) - h(\theta_n))}_{\Delta h}/T\right), 1\right)$$

Otherwise set $\theta_{n+1} = \theta_n$

④ Decrease T.

If $\Delta h < 0$, we always go

If $\Delta h > 0$, we go w/prob $\exp(-\Delta h/T)$

TOPIC:

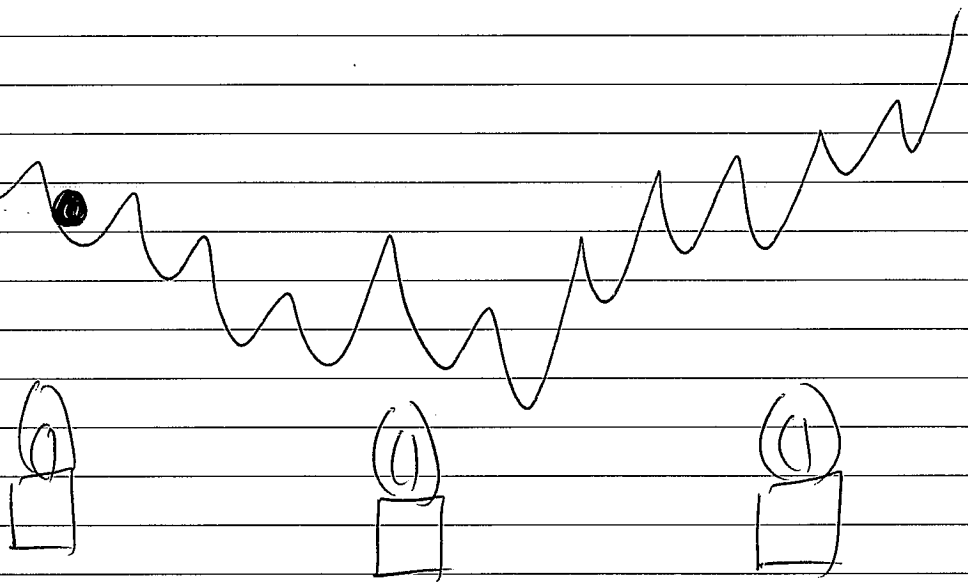
DATE:

FILE UNDER:

PAGE:

Hi temp
↳ ball moves all over

low temp
↳ ball stays in minimum



down hill

up hill

If $\Delta h < 0$, we always go

If $\Delta h > 0$, we go w/prob $\exp(-\Delta h/T)$

When T is closer to 0, it is less likely we will go up hill.

Can't cool too fast or will get stuck.

Need a cooling schedule like

$$T_n = \frac{a}{\log(n+b)}$$ (for global convergence)

Anything else is "too fast" and will get stuck in a local mode.

Thm: If

(1) $T_{n+1} < T_n \quad \forall n$

(2) $T_n \to 0$ as $n \to \infty$

(3) $T_n = a/\log(n+b)$

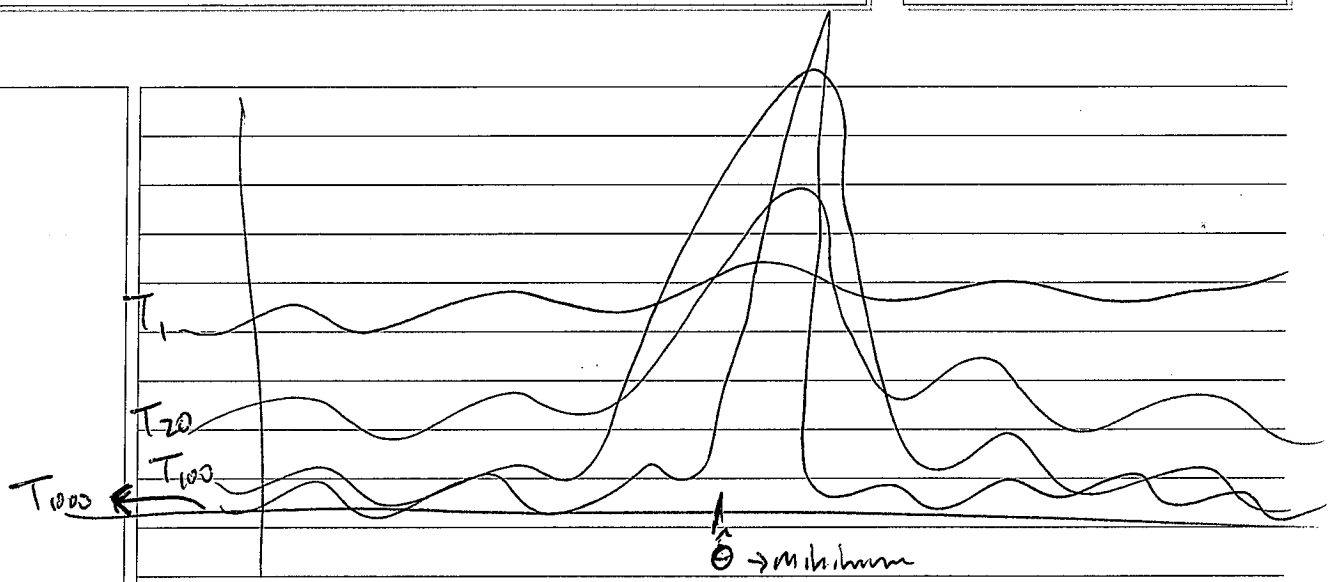
Then $\| \pi_{T_n} - \pi^* \| \to 0$

TOPIC:

DATE:

FILE UNDER:

PAGE:



As $T_n \to 0$, the mass of Π_{T_n} becomes increasingly concentrated about $\hat{\theta}$

Perfect Sampling (brief intro)

We know that under certain conditions, a MC will converge to a stationary distribution.
↳ at time $n = \infty$
But we don't know when!
We know a burn-in is a good idea, but how much?

When can we declare "convergence"?

Propp + Wilson (1996) discovered that perfect random samples can be obtained in finite (but stochastic) time.

⇒ Coupling from the Past algorithm

TOPIC:

DATE:

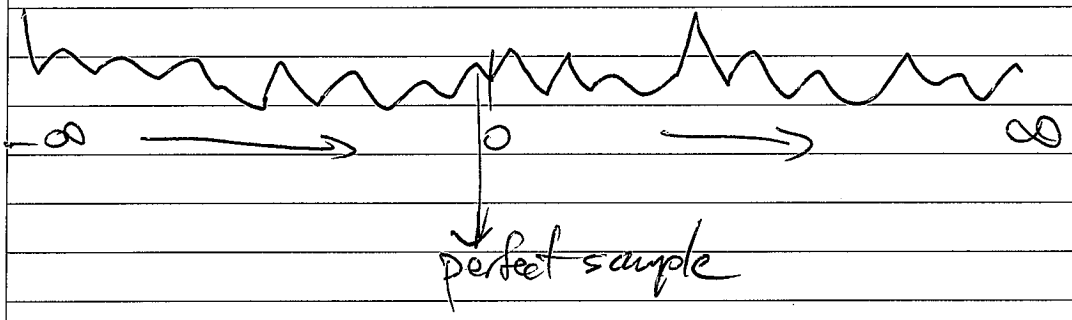
FILE UNDER:

PAGE:

Basic idea of CFTP

(1) If we start at $n=0$, then a sample from state $n=\infty$ will be a "perfect" sample from π, the stationary distribution.

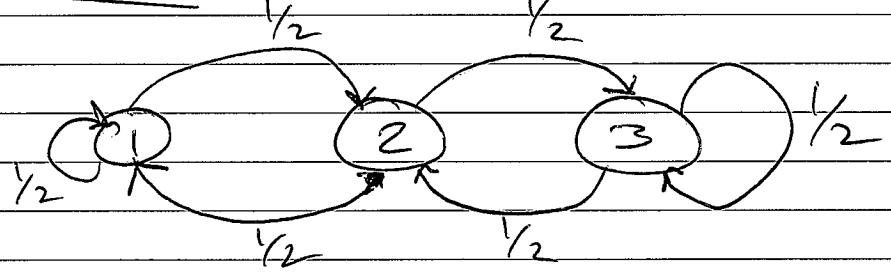
(2) So why don't we start the chain at state $n=-\infty$ and ~~take a sample from~~ use the value drawn at state $n=0$?



Both chains run for ∞ amount of time so must be stationary.

Discrete state space

Simple example



Transition matrix

$$P = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1/2 & 1/2 \end{bmatrix}$$

Stationary dist: $\pi = (1/3, 1/3, 1/3)$

TOPIC:

DATE:

FILE UNDER:

PAGE:

The stationary dist. is uniform on all 3 states.

How do we simulate this Markov Chain?

Suppose X_n is the value at state n.

Draw a $U \sim \text{uniform}(0,1)$.

If $X_n = 1$ then goto $\begin{cases} X_{n+1} = 1 & U \le \frac{1}{2} \\ X_{n+1} = 2 & U > \frac{1}{2} \end{cases}$

If $X_n = 2$ then goto $\begin{cases} X_{n+1} = 1 & U \le 1/2 \\ X_{n+1} = 3 & U > 1/2 \end{cases}$

If $X_n = 3$ then goto $\begin{cases} X_{n+1} = 2 & U \le 1/2 \\ X_{n+1} = 3 & U > 1/2 \end{cases}$

Imagine running the chain this way starting at $n = -\infty$ and ending at $n = 0$.

Can we emulate this process?

Notice:

(1) We can generate $U_n, U_{n+1}, U_{n+2}, \ldots$ because they are independent of everything.

(2) $X_{n+1} = \phi(U_n, X_n)$

TOPIC:

DATE:

FILE UNDER:

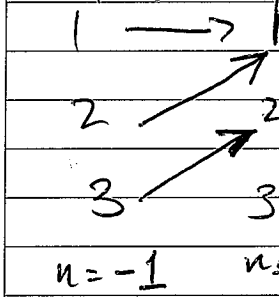
PAGE:

CFTP Algorithm for 1 perfect sample

~~Suppose we draw $U \leq 1/2$~~

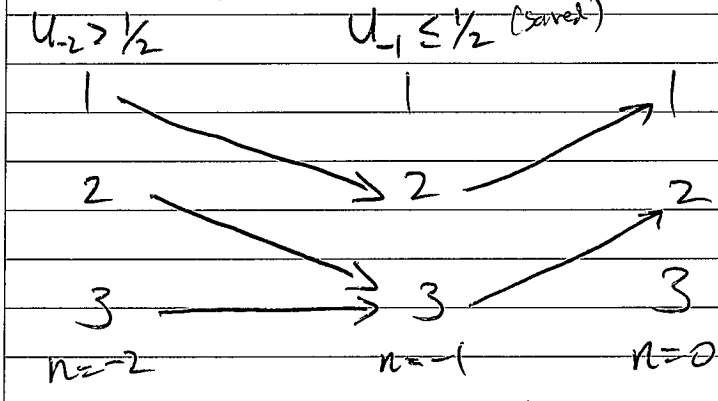
Start 3 parallel MCs each starting in state 1, 2, or 3 respectively.

Suppose we draw $U_{-1} \leq 1/2$ and start at $n=-1$



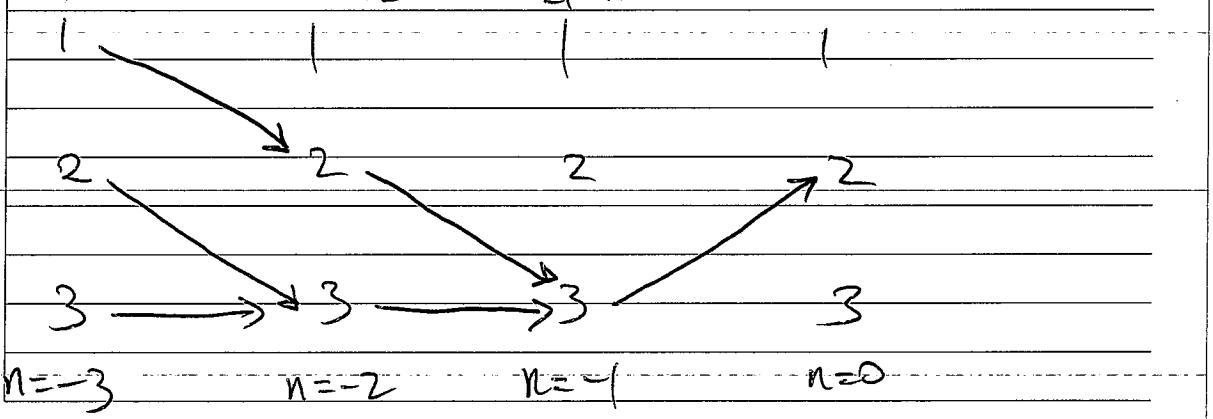
If we start the chains at $n=-1$ and draw $U \leq 1/2$, then we are at state 1 or 2. (not 3)

Suppose we generate $U_{-2} > 1/2$ and start at $n=-2$.



We "lost" the second chain, or the 2nd and 3rd chains coalesced.

Now suppose generate $U_{-3} > 1/2$ and start at $n=-3$.



TOPIC:

DATE:

FILE UNDER:

PAGE:

When we start from $n=-3$, all of the chains coalesce at state $n=-1$.

$\Rightarrow$ Regardless of what happens b/w time $n=-\infty$ and $n=-3$ (we don't know), we know that at time $n=-1$ the will all coalesce at state (3).

$\Rightarrow$ Regardless of starting value, at time $n=0$, we will be at state (2).

$\Rightarrow$ (2) is a perfect sample from stationary dist.

$\Rightarrow$ Coalescence "breaks" dependence on the starting value

Note: In this chain, the arrows don't cross, so we only need to keep track of "top" chain and "bottom" chain

(1) Why can't run forward & until coupling?
↳ only couples at (1) and (3)

(2) Why can't run until coupling and then a few more?
How many more?