Markov Chain Monte Carlo

Prof. David Page

transcribed by Matthew G. Lee

Markov Chain

- A Markov chain includes
  - A set of states
  - A set of associated transition probabilities
    - For every pair of states *s* and *s'* (not necessarily distinct) we have an associated transition probability T(*s*➔*s'*) of moving from state *s* to state *s'*
    - For any time *t*, T(*s*➔*s'*) is the probability of the Markov process being in state *s'* at time *t+1* given that it is in state *s* at time *t*

Some Properties of Markov Chains

(Some we'll use, some you may hear used elsewhere and want to know about)

- **Irreducible** chain: can get from any state to any other eventually (non-zero probability)
- **Periodic** state: state *i* is periodic with period *k* if all returns to *i* must occur in multiples of *k*
- **Ergodic** chain: irreducible and has an aperiodic state. Implies all states are aperiodic, so chain is aperiodic.
- Finite state space: can represent chain as matrix of transition probabilities… then *ergodic* = *regular*…
- **Regular** chain: some power of chain has only positive elements
- **Reversible** chain: satisfies detailed balance (**later**)

Sufficient Condition for Regularity

- A Markov chain is regular if the following properties both hold:

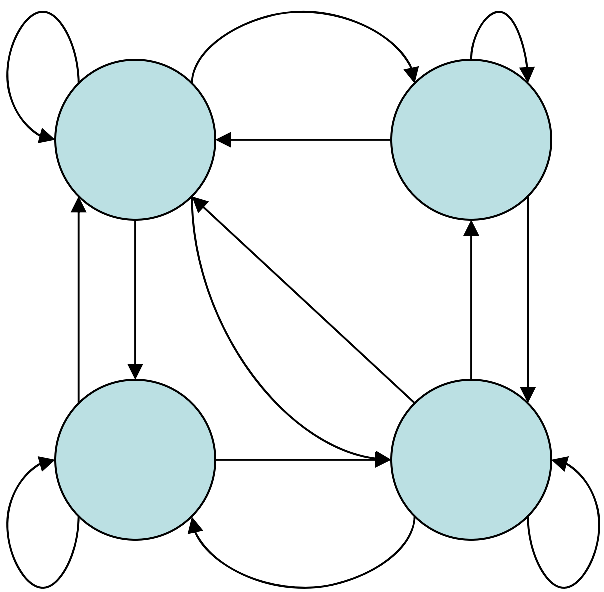
  1. For any pair of states *s*, *s'* that each have nonzero probability there exists some path from *s* to *s'* with nonzero probability

  2. For all *s* with nonzero probability, the "self loop" probability T(*s*→*s*) is nonzero

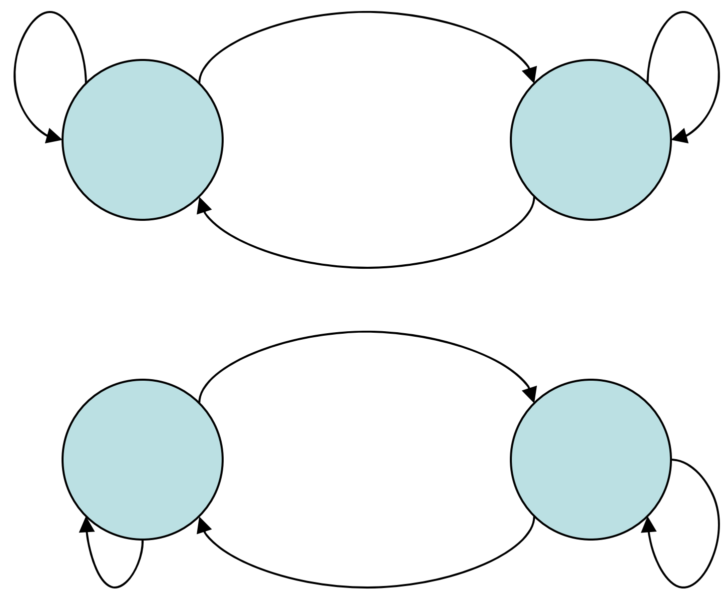
- Gibbs sampling is regular if no zeroes in CPTs

Examples of Markov Chains (arrows denote nonzero-probability transitions)

- Regular



- Non-regular

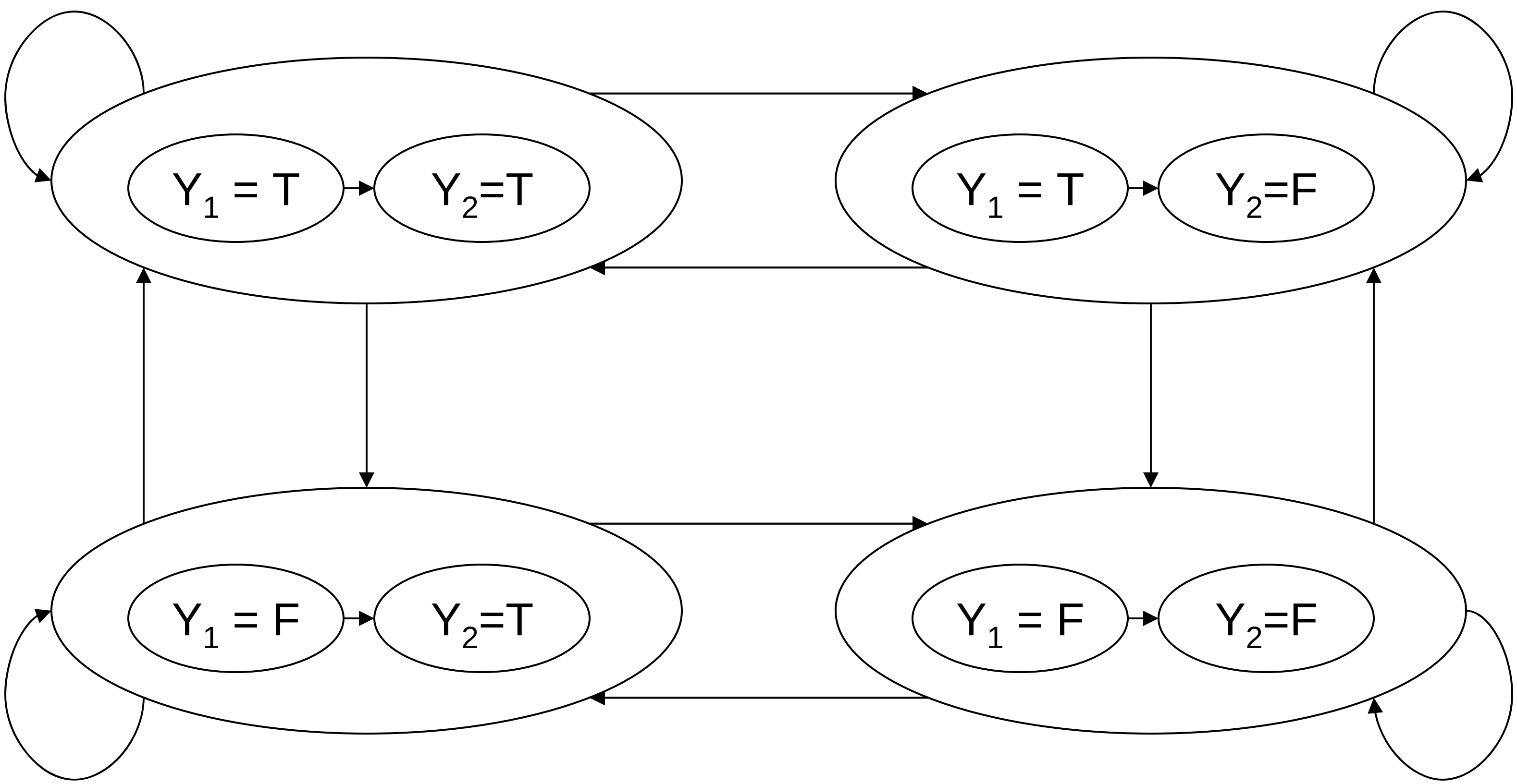


Sampling of Random Variables Defines a Markov Chain

- A state in the Markov chain is an assignment of values to all random variables

Example

- Each of the four large ovals is a state
- Transitions correspond to a Gibbs sampler

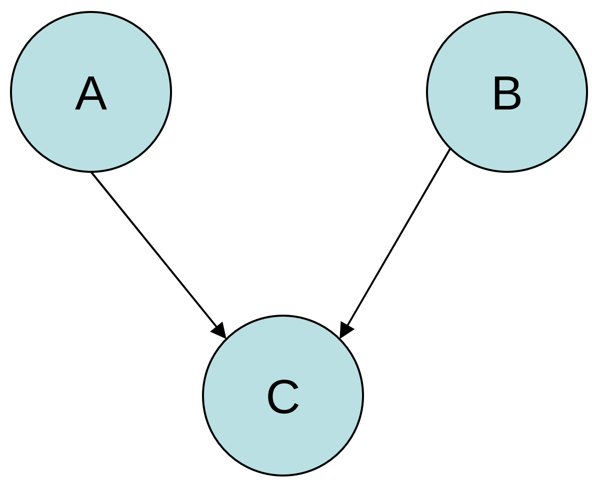


Bayes Net for which Gibbs Sampling is a Non-Regular Markov Chain

The Markov chain defined by Gibbs sampling has eight states, each an assignment to the three Boolean states A, B, and C. It is impossible to go from the state A=T, B=T, C=F to any other state

P(A)
.xx...

P(B)
.yy...



A	B	P(C)
T	T	0
T	F	1
F	T	1
F	F	0

Notation: States

- y_i and y_i' denote assignments of values to the random variable Y_i
- We abbreviate $Y_i=y_i$ by y_i
- $\boldsymbol{y}$ denotes the state of assignments $\boldsymbol{y}=(y_1,y_2,\ldots,y_n)$
- $\boldsymbol{u_i}$ is the partial description of a state given by $Y_j=y_j$ for all j not equal to i, or $(y_1,y_2,\ldots,y_{i-1},y_{i+1}\ldots,y_n)$
- Similarly, $\boldsymbol{y'}=(y_1',y_2',\ldots,y_n')$ and $\boldsymbol{u_i'}=(y_1',y_2',\ldots,y_{i-1}',y_{i+1}'\ldots,y_n')$

Notation: Probabilities

- $\pi_t(\boldsymbol{y})$ = probability of being in state ***y*** at time *t*
- Transition function $T(\boldsymbol{y} \rightarrow \boldsymbol{y}')$ = probability of moving from state ***y*** to state ***y'***

Bayesian Network Probabilities

- We use P to denote probabilities according to our Bayesian network, conditioned on the evidence
  - For example, $P(y_i'|\boldsymbol{u_i})$ is the probability that random variable Y_i has value y_i' given that $Y_j=y_j$ for all *j* not equal to *i*

Assumption: CPTs nonzero

- We will assume that all probabilities in all conditional probability tables are nonzero

- So, for any ***y***, $P(y) = \prod_{i=1}^{n} P(y_i \mid y_j \forall j \in Parents[i]) > 0$

- So, for any event *S*, $P(S) = \sum_{y \in S} P(y) > 0$

- So, for any events S_1 and S_2, $P(S_1 \mid S_2) = \frac{P(S_1 \cap S_2)}{P(S_2)} > 0$

Gibbs Sampler Markov Chain

- We assume we have already chosen to sample variable Y_i
  - $\mathrm{T}(\boldsymbol{u_i}, y_i \Rightarrow \boldsymbol{u_i}, y_i') = \mathrm{P}(y_i' | \boldsymbol{u_i})$
- If we want to incorporate the probability of randomly uniformly choosing a variable to sample, simply multiply all transition probabilities by $1/n$

Gibbs Sampler Markov Chain is Regular

- Path from ***y*** to ***y'*** with Nonzero Probability:
  - Let n be the number of variables in the Bayes net.
  - For step $i = 1$ to n :
  - Set variable Y_i to y_i' and leave other variables the same. That is, go from $(y_1',y_2',\ldots,y_{i-1}',y_i,y_{i+1},\ldots,y_n)$ to $(y_1',y_2',\ldots,y_{i-1}',y_i',y_{i+1},\ldots,y_n)$
  - The probability of this step is $P(y_i'|y_1',y_2',\ldots,y_{i-1}',y_{i+1},\ldots,y_n)$, which is nonzero
- So all steps, and thus the path, has nonzero probability
- Self loop $T(\mathbf{y} \rightarrow \mathbf{y})$ has probability $P(y_i|\mathbf{u_i}) > 0$

How π Changes with Time in a Markov Chain

- $\pi_{t+1}(\boldsymbol{y}') = \sum_{\mathbf{y}} \pi_t(\mathbf{y}) T(\mathbf{y} \rightarrow \mathbf{y}')$
- A distribution π_t is stationary if $\pi_t = \pi_{t+1}$, that is, for all $\boldsymbol{y}$, $\pi_t(\boldsymbol{y}) = \pi_{t+1}(\boldsymbol{y})$

Detailed Balance

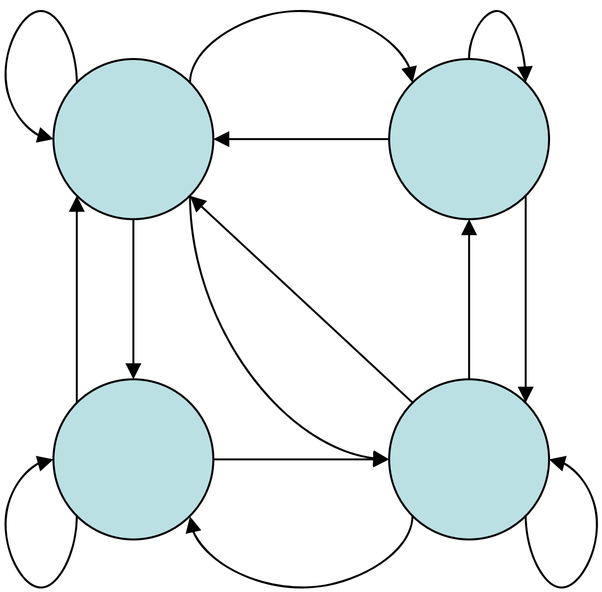
- A Markov chain satisfies detailed balance if there exists a unique distribution π such that for all states ***y***, ***y'***,

$$\pi(\boldsymbol{y})T(\boldsymbol{y} \rightarrow \boldsymbol{y'}) = \pi(\boldsymbol{y'})T(\boldsymbol{y'} \rightarrow \boldsymbol{y})$$

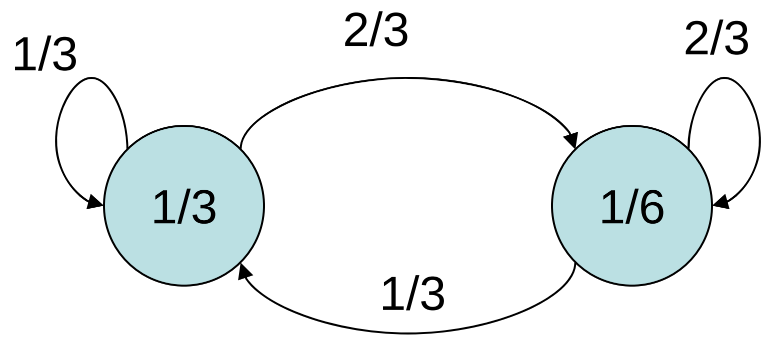
- If a regular Markov chain satisfies detailed balance with distribution π, then there exists *t* such that for any initial distribution π_0, $\pi_t = \pi$
- Detailed balance (with regularity) implies convergence to unique stationary distribution

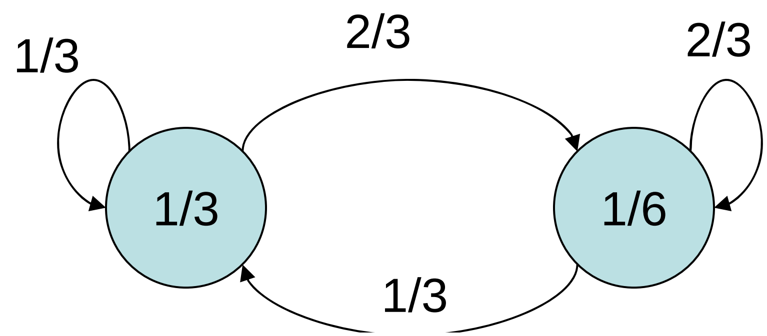
Examples of Markov Chains (arrows denote nonzero-probability transitions)

- Regular, Detailed Balance (with appropriate π and T) ➔ Converges to Stationary Distribution



- Detailed Balance with π on nodes and T on arcs. Does not converge because not regular





Gibbs Sampler satisfies Detailed Balance

Claim: A Gibbs sampler Markov chain defined by a Bayesian network with all CPT entries nonzero satisfies detailed balance with probability distribution $\pi(\boldsymbol{y})=P(\boldsymbol{y})$ for all states $\boldsymbol{y}$

Proof: First we will show that $P(\boldsymbol{y})T(\boldsymbol{y} \rightarrow \boldsymbol{y'}) = P(\boldsymbol{y'})T(\boldsymbol{y'} \rightarrow \boldsymbol{y})$. Then we will show that no other probability distribution π satisfies $\pi(\boldsymbol{y})T(\boldsymbol{y} \rightarrow \boldsymbol{y'}) = \pi(\boldsymbol{y'})T(\boldsymbol{y'} \rightarrow \boldsymbol{y})$

Gibbs Sampler satisfies Detailed Balance, Part 1

$P(\boldsymbol{y})T(\boldsymbol{y} \rightarrow \boldsymbol{y'})$	$= P(y_i, \boldsymbol{u_i})P(y_i' \mid \boldsymbol{u_i})$	(Gibbs Sampler Def.)
	$= P(y_i \mid \boldsymbol{u_i})P(\boldsymbol{u_i})P(y_i' \mid \boldsymbol{u_i})$	(Chain Rule)
	$= P(y_i', \boldsymbol{u_i})P(y_i \mid \boldsymbol{u_i})$	(Reverse Chain Rule)
	$= P(\boldsymbol{y'})T(\boldsymbol{y'} \rightarrow \boldsymbol{y})$	(Gibbs Sampler Def.)

Gibbs Sampler Satisfies Detailed Balance, Part 2

Since all CPT entries are nonzero, the Markov chain is regular. Suppose there exists a probability distribution π not equal to P such that $\pi(\boldsymbol{y})T(\boldsymbol{y} \rightarrow \boldsymbol{y'}) = \pi(\boldsymbol{y'})T(\boldsymbol{y'} \rightarrow \boldsymbol{y})$. Without loss of generality, there exists some state $\boldsymbol{y}$ such that $\pi(\boldsymbol{y}) > P(\boldsymbol{y})$. So, for every neighbor $\boldsymbol{y'}$ of $\boldsymbol{y}$, that is, every $\boldsymbol{y'}$ such that $T(\boldsymbol{y} \rightarrow \boldsymbol{y'})$ is nonzero,

$$\pi(\boldsymbol{y'})T(\boldsymbol{y'} \rightarrow \boldsymbol{y}) = \pi(\boldsymbol{y})T(\boldsymbol{y} \rightarrow \boldsymbol{y'}) > P(\boldsymbol{y})T(\boldsymbol{y} \rightarrow \boldsymbol{y'}) = P(\boldsymbol{y'})T(\boldsymbol{y'} \rightarrow \boldsymbol{y})$$

So $\pi(\boldsymbol{y'}) > P(\boldsymbol{y'})$.

Gibbs Sampler Satisfies Detailed Balance, Part 3

We can inductively see that $\pi(\boldsymbol{y}'') > P(\boldsymbol{y}'')$ for every state $\boldsymbol{y}''$ path-reachable from $\boldsymbol{y}$ with nonzero probability. Since the Markov chain is regular, $\pi(\boldsymbol{y}'') > P(\boldsymbol{y}'')$ for all states $\boldsymbol{y}''$ with nonzero probability. But the sum over all states $\boldsymbol{y}''$ of $\pi(\boldsymbol{y}'')$ is 1, and the sum over all states $\boldsymbol{y}''$ of $P(\boldsymbol{y}'')$ is 1. This is a contradiction. So we can conclude that P is the unique probability distribution π satisfying $\pi(\boldsymbol{y})T(\boldsymbol{y} \rightarrow \boldsymbol{y}') = \pi(\boldsymbol{y}')T(\boldsymbol{y}' \rightarrow \boldsymbol{y})$.

Using Other Samplers

- The Gibbs sampler only changes one random variable at a time
  - Slow convergence
  - High-probability states may not be reached because reaching them requires going through low-probability states

Metropolis Sampler

- Propose a transition with probability $T^Q(\boldsymbol{y} \rightarrow \boldsymbol{y}')$
- Accept with probability $A=\min(1, P(\boldsymbol{y}')/P(\boldsymbol{y}))$
- If for all $\boldsymbol{y}$, $\boldsymbol{y}'$ $T^Q(\boldsymbol{y} \rightarrow \boldsymbol{y}')=T^Q(\boldsymbol{y}' \rightarrow \boldsymbol{y})$ then the resulting Markov chain satisfies detailed balance

Metropolis-Hastings Sampler

- Propose a transition with probability $T^Q(\boldsymbol{y} \rightarrow \boldsymbol{y'})$
- Accept with probability

  $A=\min(1, P(\boldsymbol{y'})T^Q(\boldsymbol{y'} \rightarrow \boldsymbol{y})/P(\boldsymbol{y})T^Q(\boldsymbol{y} \rightarrow \boldsymbol{y'}))$
- Detailed balance satisfied
- Acceptance probability often easy to compute even though sampling according to P difficult

Gibbs Sampler as Instance of Metropolis-Hastings

- Proposal distribution $T^Q(\boldsymbol{u_i}, y_i \rightarrow \boldsymbol{u_i}, y_i') = P(y_i' | \boldsymbol{u_i})$
- Acceptance probability

$$A = \min(1, \frac{P(u_i, y_i')T^Q(u_i, y_i' \rightarrow u_i, y_i)}{P(u_i, y_i)T^Q(u_i, y_i \rightarrow u_i, y_i')})$$

$$= \min(1, \frac{P(y_i' | u_i)P(u_i)P(y_i | u_i)}{P(y_i | u_i)P(u_i)P(y_i' | u_i)})$$

$$= 1$$