Probabilistic Graphical Models

Lectures 20

Sampling Multivariable Distributions
Ancestral Sampling for Bayes Nets

Sampling from a multi-variable distribution

$$P(x,y) \sim (x^1,y^1), (x^2,y^2), \dots, (x^m,y^m)$$

$$p(x,y) = p(x)\, p(y|x)$$

$$p(x) = \sum_y p(x,y)$$

$$p(y|x) = \frac{p(x,y)}{p(x)}$$

$$x^i \sim p(x)$$

$$y^i \sim p(y|x^i)$$

claim: (x^i, y^i) is a sample from $p(x,y)$

proof: $\Pr(X = x^i, Y = y^i) = p(x^i)\, p(y^i|x^i) = p(x^i, y^i)$

$\Longrightarrow$ (x^i, y^i) is a sample drawn from $p(x,y)$

Sampling from a multi-variable distribution

$p(x,y) = p(x)\,p(y|x)$ (Solution 1: Chain Rule)

$x_i \sim p(x)$

$y_i \sim p(y|x_i)$

(x_i, y_i) is a sample from $p(x,y)$

Probability of the ocurranse of $(x_i, y_i) = p(x_i)\,p(y_i|x_i) = p(x_i, y_i)$

Solution 1: Using Chain Rule

$$p(x_1, x_2, \dots, x_n) = p(x_n | x_1 — x_{n-1}) \underbrace{p(x_1 — x_{n-1})}$$

25 II

$$\vdots$$

$$= p(x_n | x_1 — x_{n-1})\, p(x_{n-1} | x_1 — x_{n-2}) \cdots p(x_2 | x_1)\, p(x_1)$$

Sampling using chain rule

$x_1^i \sim p(x_1)$

$x_2^i \sim p(x_2 | x_1^i)$

$x_3^i \sim p(x_3 | x_1^i, x_2^i)$

$\vdots$

$x_n^i \sim p(x_n | x_1^i, x_2^i \cdots x_{n-1}^i)$

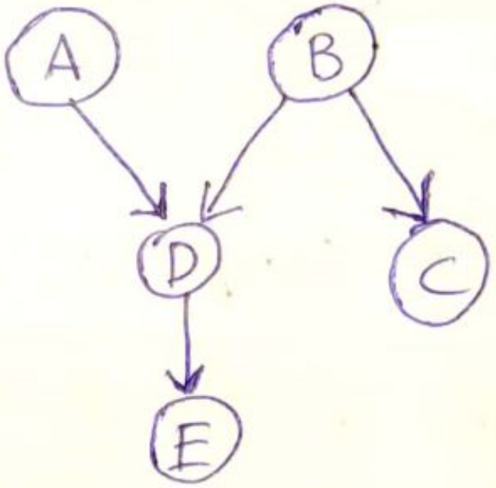
$\Longrightarrow (x_1^i, x_2^i, \dots, x_n^i)$ is a sample from $p(x_1, — , x_n)$

Sampling from a Bayesian Network

Sampling from a Bayes Net

$$P(A,B,C,D,E) = P(E \mid A,B,C,D)\, P(A,B,C,D)$$
$$= P(E \mid D)\, P(A,B,C,D)$$

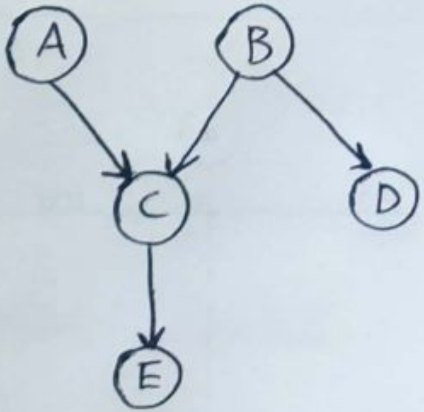
$$P(A,B,C,D,E) = P(A)\, P(B)\, P(C \mid B)\, P(D \mid A,B)\, P(E \mid D)$$



Sampling from a Bayesian Network

$$P(A,B,C,D,E)$$

$$= P(A)\,P(B)\,P(C|A,B)\,P(D|B)\,P(E|C)$$



$$a^i \sim P(A)$$

$$b^i \sim P(B)$$

$$c^i \sim P(C|a^i, b^i)$$

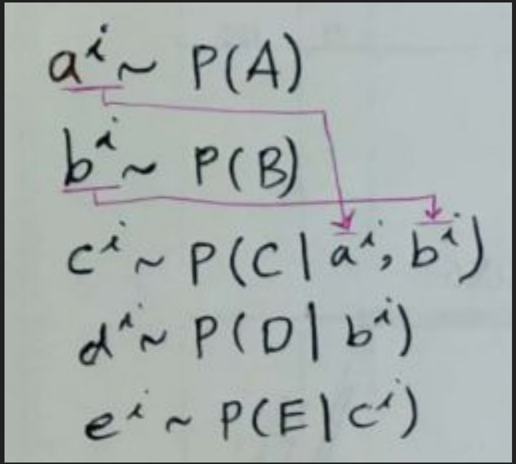
$$d^i \sim P(D|b^i)$$

$$e^i \sim P(E|c^i)$$

$(a^i, b^i, c^i, d^i, e^i)$ is a sample from $P(A,B,C,D,E)$

Ancestral Sampling

$$P(A,B,C,D,E) = P(A)\,P(B)\,P(C|B)\,P(D|A,B)\,P(E|D)$$



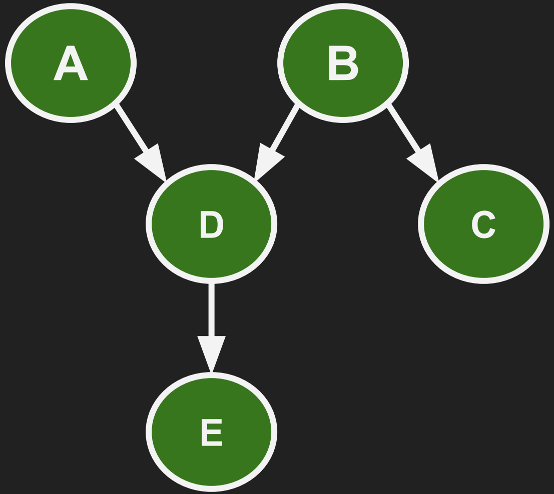
$a^i \sim P(A)$

$b^i \sim P(B)$

$c^i \sim P(C \mid a^i, b^i)$

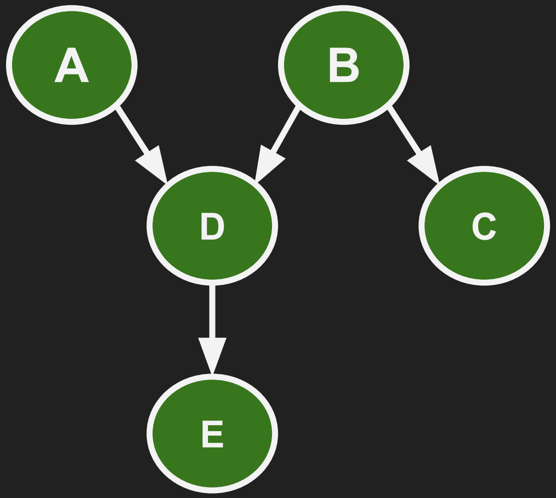
$d^i \sim P(D \mid b^i)$

$e^i \sim P(E \mid c^i)$



$(a^i, b^i, c^i, d^i, e^i)$ is a sample from $P(A,B,C,D,E)$

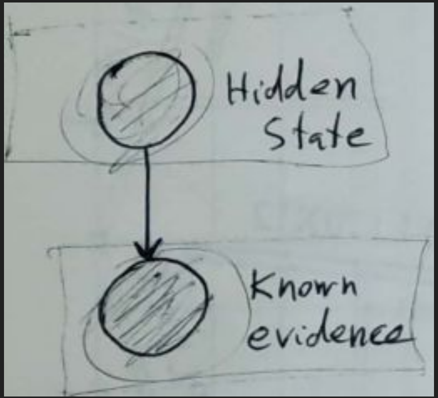
Inference using samples

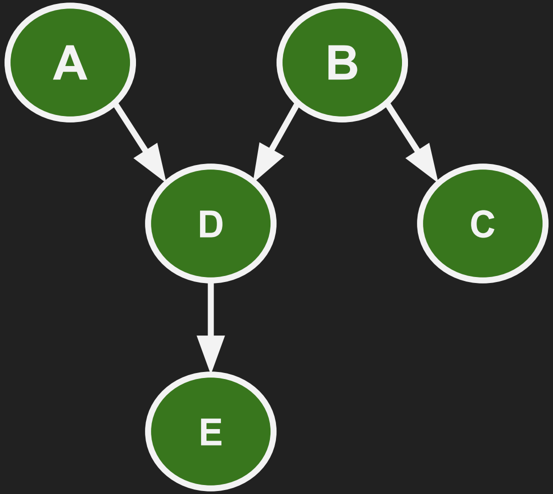


Samples: (A^1,B^1,C^1,D^1,E^1), (A^2,B^2,C^2,D^2,E^2), ..., (A^m,B^m,C^m,D^m,E^m)

$P(E=0) \approx \#(E^i = 0) / m$

Inference using samples

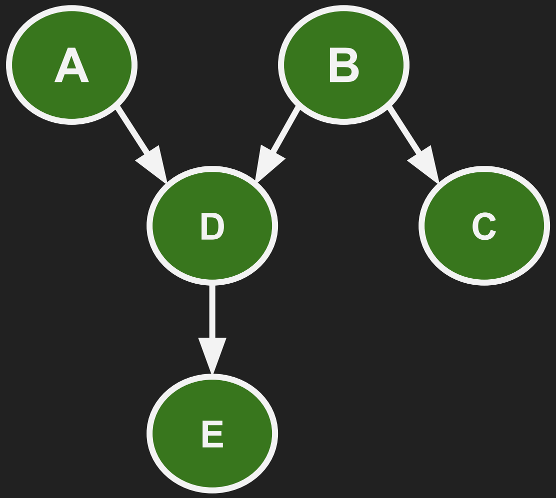




Samples: (A^1,B^1,C^1,D^1,E^1), (A^2,B^2,C^2,D^2,E^2), ..., (A^m,B^m,C^m,D^m,E^m)

$P(E=0) \approx \#(E^i = 0) / m$

Inference with evidence



Samples: $(A^1,B^1,C^1,D^1,E^1), (A^2,B^2,C^2,D^2,E^2), \ldots, (A^m,B^m,C^m,D^m,E^m)$

$P(D=0 \mid E=1) \approx \#(D^i = 0, E^i = 1) \,/\, \#(E^i = 1)$

Inference with evidence

Samples: (A^1,B^1,C^1,D^1,E^1), (A^2,B^2,C^2,D^2,E^2), ..., (A^m,B^m,C^m,D^m,E^m)

$$P(D=1 \mid A=2, B=0, C=2, E=3)$$

$$\frac{\#(D^i=1, A^i=2, B^i=0, C^i=2, E^i=3)}{\#(A^i=2, B^i=0, C^i=2, E^i=3)}$$

Inference with evidence

Samples: $(A^1,B^1,C^1,D^1,E^1), (A^2,B^2,C^2,D^2,E^2), \ldots, (A^m,B^m,C^m,D^m,E^m)$

$$P(D=1 \mid A=2, B=0, C=2, E=3) = \frac{\#(D^i=1, A^i=2, B^i=0, C^i=2, E^i=3)}{\#(A^i=2, B^i=0, C^i=2, E^i=3)}$$

Need exponentially large number of samples as the number of query variables grow.

Sampling challenges

- How to efficiently sample from a Bayes Net with evidence?
- How to sample from a Markov Random Field?
  - Can we use chain rule?

Bayes Net with evidence as an MRF

K. N. Toosi
University of Technology

$$P(A,B,C,D,E) = P(A)\,P(B)\,P(C|A,B)\,P(D|B)\,P(E|C)$$

$\boxed{\begin{array}{c} P(A|D,E) \\ P(A|D=d, E=e) \end{array}}$ $\Rightarrow$ solution 1: samples from $P(A,B,C,D,E)$

solution 2: samples from $p(A,B,C|d,e)$

$$P(A,B,C|d,e) = \frac{P(A)\,P(B)\,P(C|A,B)\,p(d|B)\,p(e|C)}{\sum_A \sum_B \sum_C P(A)\,p(B)\,P(C|A,B)\,P(d|B)\,p(e|C)}$$

$$= \frac{1}{Z} P(A)\,P(B)\,P(C|A,B)\,P(d|B)\,P(e|C)$$

$$= \frac{1}{Z} \Phi_1(A)\,\Phi_2(B)\,\Phi_3(A,B,C)\,\Phi_4(B)\,\Phi_5(C)$$

Bayes Net with evidence as an MRF

$$P(A,B,C,D,E) = P(A)\,P(B)\,P(C|A,B)\,p(D|B)\,P(E|C)$$

$\boxed{P(A|D,E)}$ $\Rightarrow$ solution 1: samples from $P(A,B,C,D,E)$

$\boxed{P(A|D=d, E=e)}$ solution 2: samples from $p(A,B,C|d,e)$

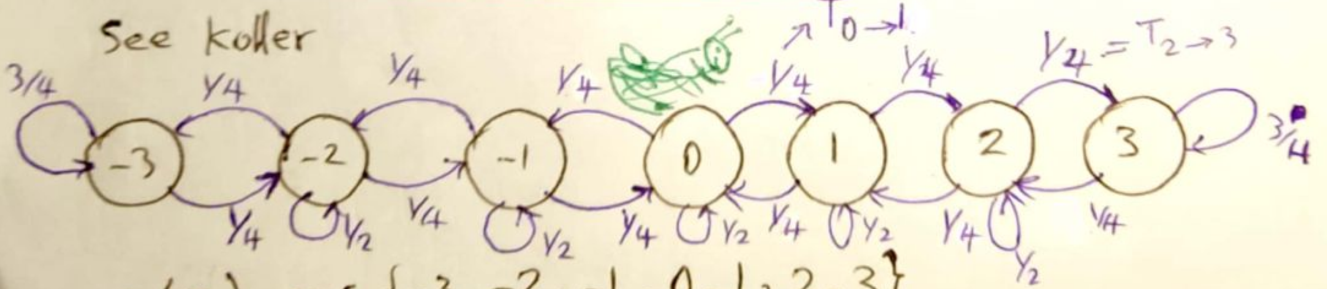
$$P(A,B,C|d,e) = \frac{P(A)\,P(B)\,P(C|A,B)\,p(d|B)\,p(e|C)}{\sum_A \sum_B \sum_C P(A)\,p(B)\,P(C|A,B)\,P(d|B)\,p(e|c)}$$

$$= \frac{1}{Z} P(A)\,P(B)\,P(C|A,B)\,P(d|B)\,P(e|c)$$

$$= \frac{1}{Z}\,\phi_1(A)\,\phi_2(B)\,\phi_3(A,B,C)\,\phi_4(B)\,\phi_5(C)$$

It boils down to sampling from an MRF

Sampling Using a Markov Chain



$p(x) \quad x \in \{-3, -2, -1, 0, 1, 2, 3\}$

$T(\quad T_{m \to n} = p(S_{t+1} = n \mid S_t = m)$ منتقل از t

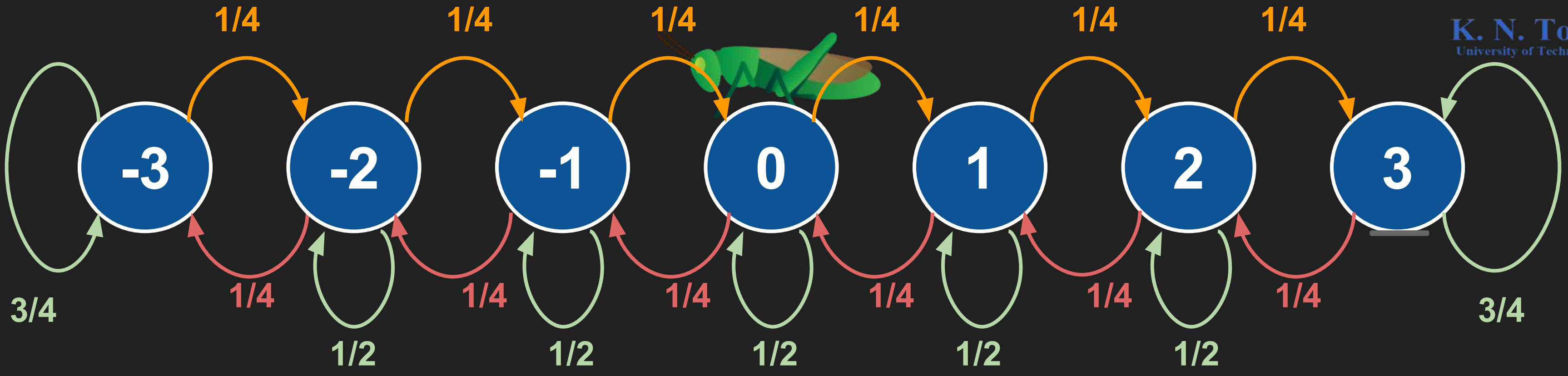
$T_{0\to 0}$

	-3	-2	-1	0	1	2	3
$t=0$	0	0	0	1	0	0	0
$t=1$	0	0	1/4	1/2	1/4	0	0
$t=2$	0	1/16	1/4	3/8	1/4	1/16	0

$1/4 * 1/2 + 1/2 * 1/4$

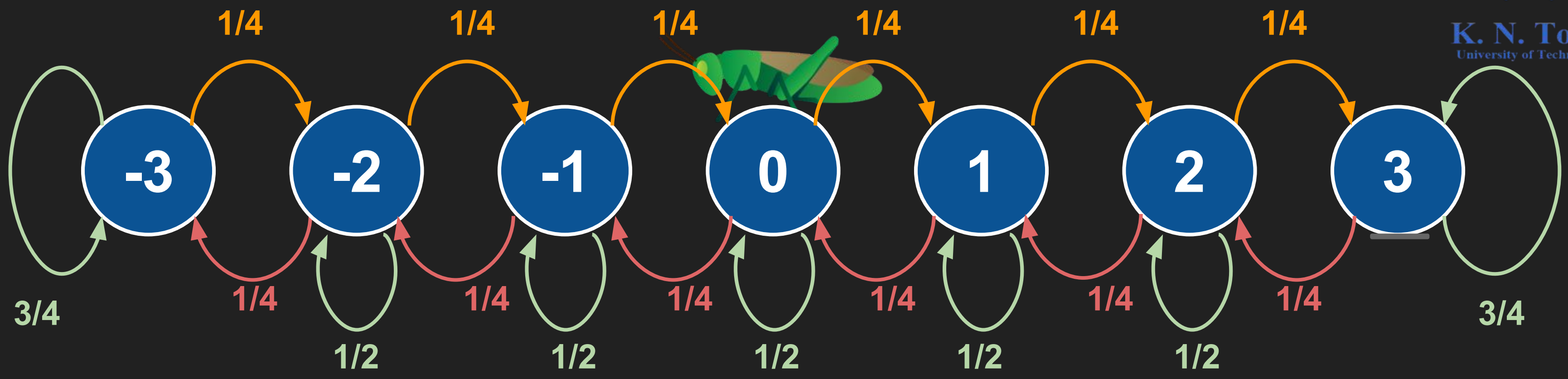
$1/2 * \frac{1}{2} + \frac{1}{4} * 1/4 + 1/4 * 1/4$

Sampling Using a Markov Chain (Koller)



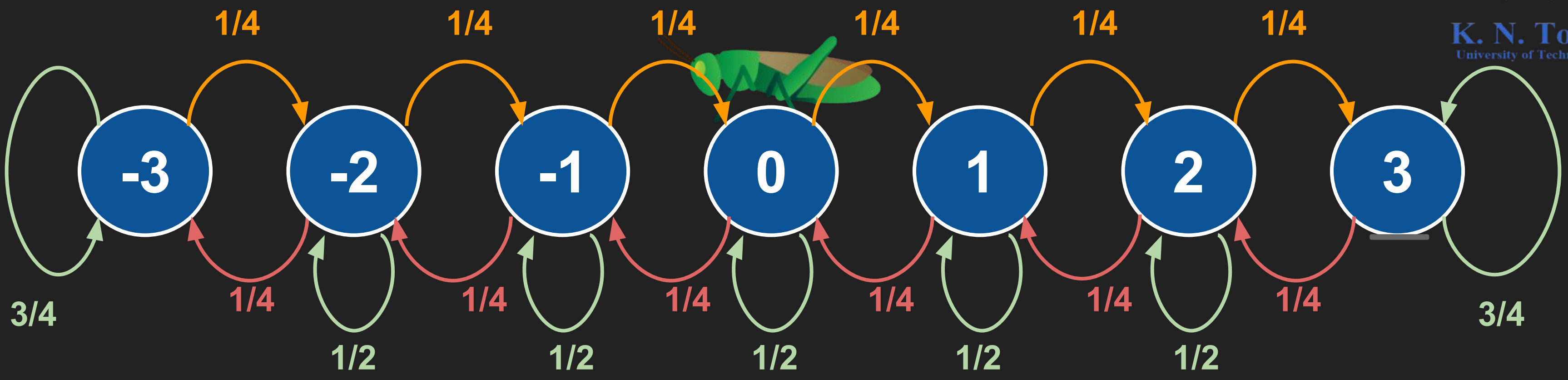
	-3	-2	-1	0	1	2	3
t=0	0	0	0	1	0	0	0
t=1							
t=2							

Sampling Using a Markov Chain (Koller)



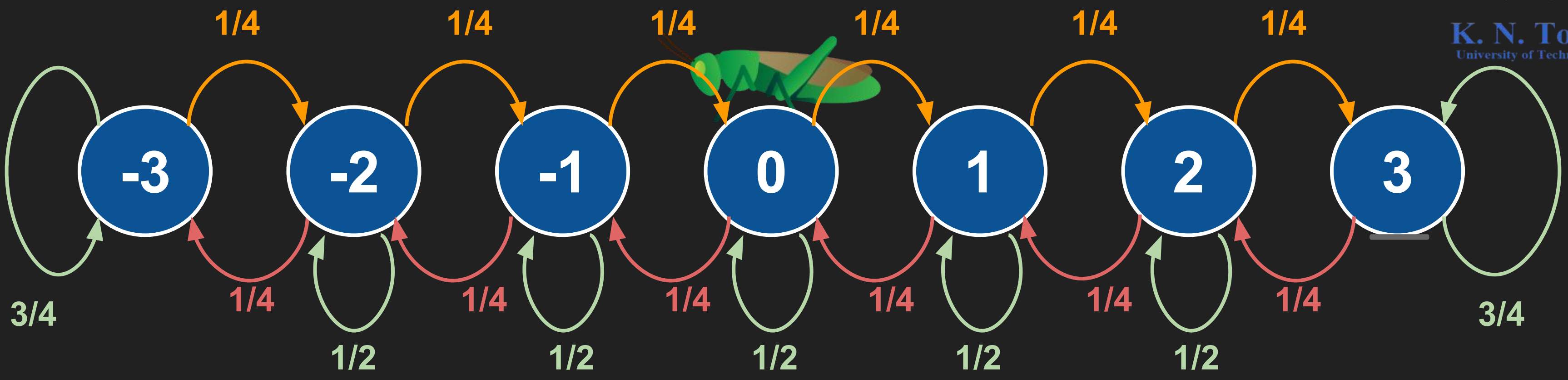
	-3	-2	-1	0	1	2	3
t=0	0	0	0	1	0	0	0
t=1	0	0	1/4	1/2	1/4	0	0
t=2							

Sampling Using a Markov Chain (Koller)



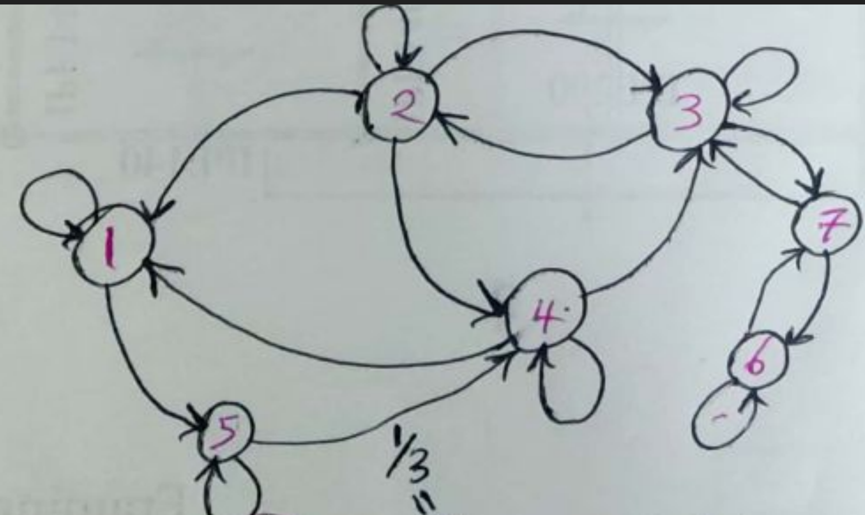
	-3	-2	-1	0	1	2	3
t=0	0	0	0	1	0	0	0
t=1	0	0	1/4	1/2	1/4	0	0
t=2	0	1/16	1/4	3/8	1/4	1/16	0

Sampling Using a Markov Chain (Koller)



	-3	-2	-1	0	1	2	3
t=0	0	0	0	1	0	0	0
:	:	:	:	:	:	:	:
$t \to \infty$	1/7	1/7	1/7	1/7	1/7	1/7	1/7

State Transition Graph



$P(X)$

$X \in \{1,2,3,4,5,6,7\}$

X_t = ~~stat~~ state ~~curren~~ at time t

$\frac{1}{3}$

$\frac{2}{3}$

$T_{5\to5}$

$$\frac{1}{3} = T_{5\to4} = Pr(X_{t+1}=4 \mid X_t=5)$$

$$= Pr(X_t=4 \mid X_{t-1}=5)$$

independent of t

$$Pr(X_t=a) = \sum_{b=1}^{7} Pr(X_t = a \mid X_{t-1}=b)\, p(X_t=b)$$