MCMC

ESC 6^{th} week

김정규

MCMC

The big picture

Monte Carlo method

Topic1: Posterior inference of arbitrary functions

Topic2: Predictive distributions and model checking

Topic3: Rejection sampling, importance sampling

MC vs. MCMC

Recap: The big picture

1. Exact Bayesian Inference (~Week5)

    Conjugacy

2. Approximate Bayesian inference (Week6 ~)

1) Marginalization

MC (Independent Monte Carlo)

MCMC (Markov Chain Monte Carlo)

Metropolis Hastings

Gibbs sampling

2) Optimization

Variational Inference

Monte Carlo method

Motivation: how to calculate?

$$\Pr(\theta_1 > \theta_2 | \sum Y_{i,1} = 217, \sum Y_{i,2} = 66)$$

$$p(\theta_1 | \sum_{i=1}^{111} Y_{i,1} = 217) = \text{dgamma}(\theta_1, 219, 112)$$

$$p(\theta_2 | \sum_{i=1}^{44} Y_{i,2} = 66) = \text{dgamma}(\theta_2, 68, 45)$$

Paper and pencil (Analytic)

$$\begin{aligned}
&\Pr(\theta_1 > \theta_2 | y_{1,1}, \ldots, y_{n_2,2}) \\
&= \int_0^\infty \int_0^{\theta_1} p(\theta_1, \theta_2 | y_{1,1}, \ldots, y_{n_2,2}) \, d\theta_2 d\theta_1 \\
&= \int_0^\infty \int_0^{\theta_1} \text{dgamma}(\theta_1, 219, 112) \times \text{dgamma}(\theta_2, 68, 45) \, d\theta_2 d\theta_1 \\
&= \frac{112^{219} 45^{68}}{\Gamma(219)\Gamma(68)} \int_0^\infty \int_0^{\theta_1} \theta_1^{218} \theta_2^{67} e^{-112\theta_1 - 45\theta_2} \, d\theta_2 d\theta_1.
\end{aligned}$$

Simlation (Monte Carlo)

```
> a<-2 ; b<-1
> sy1<-217 ;  n1<-111
> sy2<-66  ;  n2<-44

> theta1.mc<-rgamma(10000,a+sy1 , b+n1)
> theta2.mc<-rgamma(10000,a+sy2 , b+n2)

> mean(theta1.mc>theta2.mc)

[1] 0.9708
```

Monte Carlo method

- Monte Carlo Samples
  - $\theta^{(1)}, \ldots, \theta^{(S)} \overset{iid}{\sim} p(\theta|y_1, \ldots, y_n)$

- Monte Carlo Integration (random)
  - Statistical estimation of the value of an integral using Monte Carlo samples
    - $E(g(\theta)|y_1, \ldots, y_n) = \int g(\theta)\, p(\theta|\, y_1, \ldots, y_n) d\theta \approx \frac{1}{S} \sum_{s=1}^{S} g(\theta^S)$

- Numerical Integration (non-random)
  - Deterministic approximation of integration at pre-selected points
    - $E(g(\theta)|y_1, \ldots, y_n) = \int g(\theta)\, p(\theta|\, y_1, \ldots, y_n) d\theta \approx \frac{1}{S} \sum_{s=1}^{S} w_s g\left(\theta_d^S\right) p(\theta_d^S|\, y_1, \ldots, y_n)$
    - $\theta_{\mathrm{d}}^{(\mathrm{S})}$: deterministic sample points
    - $\mathrm{w_s}$: volume of space represented by the point $\theta_{\mathrm{d}}^{(\mathrm{S})}$

Monte Carlo method

- Thm1. Consistency
  - Let $g(\theta)$ be any (computable) function
  - $\frac{1}{S}\sum_{s=1}^{S} g(\theta^{S}) \rightarrow \ E[g(\theta)\ |y_1, \dots, y_n\ ] = \int g(\theta)\, p(\theta|\ y_1, \dots, y_n)\, d\theta, as\ S \rightarrow \infty$ by LLN

Monte Carlo method

- Corollary 1

  Thm1 implies that

  - $\bar{\theta} = \sum_{s=1}^{S} \frac{\theta^{(s)}}{S} \rightarrow E[\theta | y_1, \ldots, y_n]$
  - $\sum_{s=1}^{S} \frac{(\theta^{(s)} - \bar{\theta})}{S-1} \rightarrow Var[\theta | y_1, \ldots, y_n]$
  - $\frac{\#(\theta^{(s)} \leq c)}{S} \rightarrow Pr[\theta \leq c | y_1, \ldots, y_n]$
  - $empirical\ distn\ of\ \{\theta^{(1)}, \ldots, \theta^{(S)}\} \rightarrow p(\theta | y_1, \ldots, y_n)$
  - $median\left(\{\theta^{(1)}, \ldots, \theta^{(S)}\}\right) \rightarrow \theta_{\left(\frac{1}{2}\right)}$
  - $\alpha - percentile\left(\{\theta^{(1)}, \ldots, \theta^{(S)}\}\right) \rightarrow \theta_{\alpha}$

Monte Carlo method

- Histograms (with Kernel density estimates) and True Distn

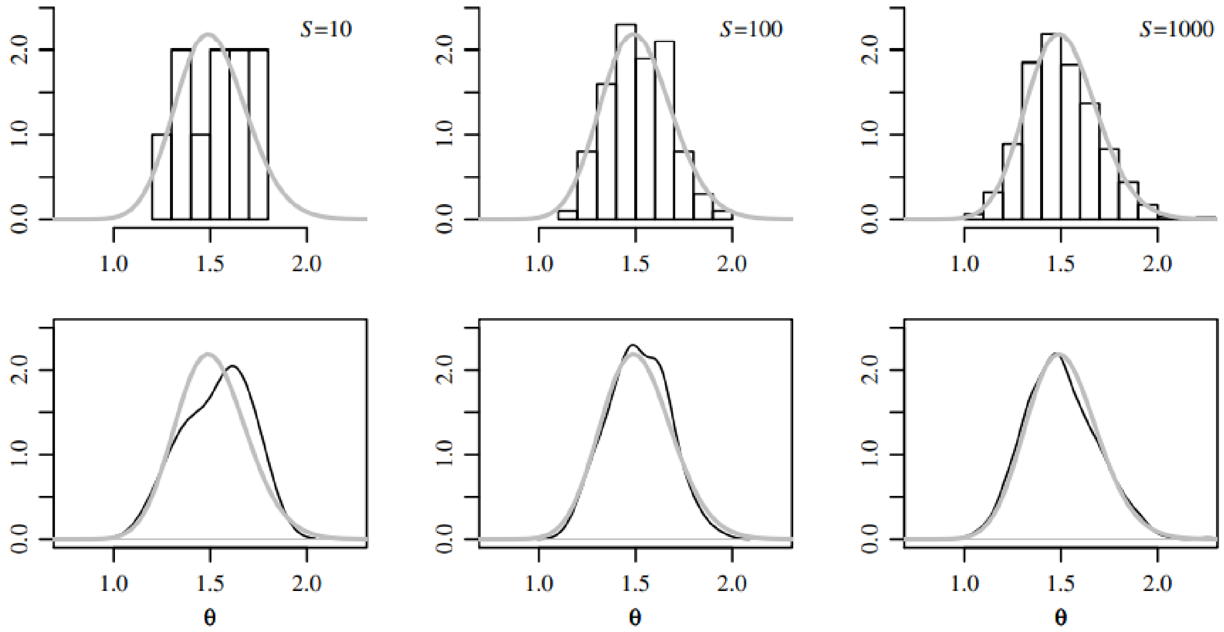
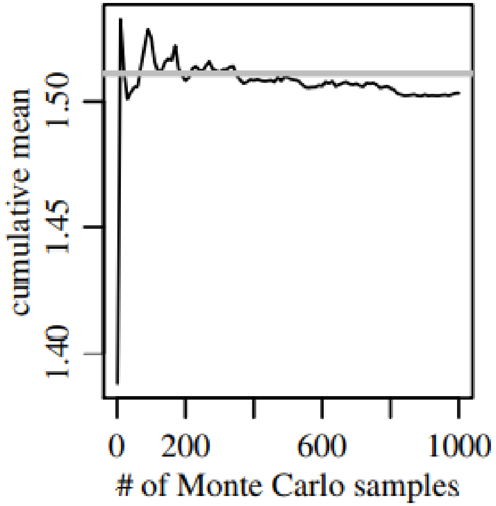


Fig. 4.1. Histograms and kernel density estimates of Monte Carlo approximations to the gamma(68,45) distribution, with the true density in gray.

Monte Carlo method

- Monte Carlo Standard Error (for posterior mean)
  - From Corollary 1,
    - $\theta \sim p(\theta|y_1, \dots y_n)$
    - Then $\bar{\theta} = \sum_{s=1}^{S} \frac{\theta^{(s)}}{S} \xrightarrow{D} N\left(E[\theta \,|y_1, \dots, y_n\,], Var[\theta|y_1, \dots, y_n]/S \,\right)$ by CLT
  - MCSE
    - $\hat{\sigma} = \sum_{s=1}^{S} \frac{(\theta^{(s)} - \bar{\theta})}{S-1} \approx \sqrt{Var[\theta|y_1, \dots, y_n]/S}$
    - $MCSE(mean) = \sqrt{\frac{\hat{\sigma}}{S}}$
  - Choose sample size S to attain desired precision



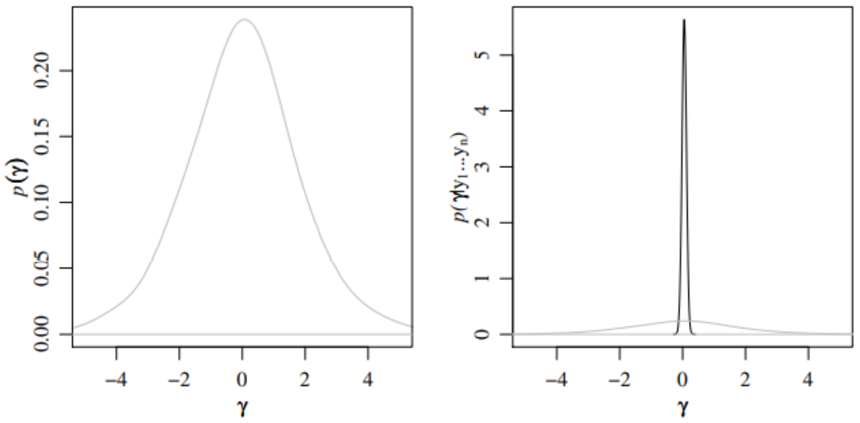
Topic1. Posterior Inference of arbitrary functions

- Ex) Log odds
  - $\mathrm{g}(\theta) = \log \mathrm{odds}(\theta) = \log \frac{\theta}{1-\theta} = \gamma$
  - Get S independent samples of $\theta^{(\mathrm{i})} \sim \mathrm{p}(\theta|\mathrm{Data}) \rightarrow \gamma^{(\mathrm{i})} = \mathrm{g}\left(\theta^{(i)}\right)$
  - $\{\gamma^{(1)}, \ldots, \gamma^{(\mathrm{S})}\}$ are S independent samples from $\mathrm{p}(\gamma|\mathrm{Data})$
  - Apply thm 1. to get $\mathrm{E}[\gamma\,|\mathrm{Data}\,]$, $\mathrm{Var}[\gamma\,|\mathrm{Data}\,]$, $\mathrm{p}(\gamma\,|\mathrm{Data}\,)$

Topic1. Posterior Inference of arbitrary functions

- Example1
  - 1998 General Social Survey
  - Agree with supreme court's ruling (prohibit requirement to read religious texts in public schools) ?
  - Out of 1011 Protestants, 353 agreed
  - Out of 860 non-Protestants (=minority), 441 agreed
  - θ: *population proportion who agrees*

```
a<-1 ; b<-1
theta.prior.mc<-rbeta(10000,a,b)
gamma.prior.mc<- log( theta.prior.mc/(1-theta.prior.mc) )

n0<-860-441 ; n1<-441
theta.post.mc<-rbeta(10000,a+n1,b+n0)
gamma.post.mc<- log( theta.post.mc/(1-theta.post.mc) )
```



Topic1. Posterior Inference of arbitrary functions

- Example2
  - Two educational groups, difference in birthrate θ
  - $\{\theta_1 \mid Data_1\} \sim gamma\ (217+2, 111+1)$: $women\ without\ bachelor's\ degree$
  - $\{\theta_2 \mid Data_2\} \sim gamma\ (66+2, 44+1)$: $women\ with\ bachelor's\ degree$
  - Using Monte Carlo, $\frac{1}{S}\sum_{s=1} I(\theta_1^s > \theta_2^s) \approx \Pr(\theta_1 > \theta_2 \mid Data)$

$$\Pr(\theta_1 > \theta_2 \mid \sum Y_{i,1} = 217, \sum Y_{i,2} = 66)$$

$$p(\theta_1 \mid \sum_{i=1}^{111} Y_{i,1} = 217) = \text{dgamma}(\theta_1, 219, 112)$$

$$p(\theta_2 \mid \sum_{i=1}^{44} Y_{i,2} = 66) = \text{dgamma}(\theta_2, 68, 45)$$

```
> a<-2 ; b<-1
> sy1<-217 ;  n1<-111
> sy2<-66  ;  n2<-44

> theta1.mc<-rgamma(10000,a+sy1, b+n1)
> theta2.mc<-rgamma(10000,a+sy2, b+n2)

> mean(theta1.mc>theta2.mc)

[1] 0.9708
```

Topic2. Predictive distributions and model checking

- Predictive model
  - θ is unknown -> integrate out
  - Prior predictive
    - $\Pr\left(\tilde{Y} = \tilde{y}\right) = \int p(\tilde{y}|\theta)\, p(\theta)\, d\theta$
  - Posterior predictive
    - $\Pr\left(\tilde{Y} = \tilde{y} \mid Data\right) = \int p(\tilde{y}|\theta, Data)\, p\,(\theta|\, Data) d\theta \; = \int p(\tilde{y}|\theta) p(\theta|Data) d\theta$

Topic2. Predictive distributions and model checking

- Predictive model
  - Sample $\theta^{(1)} \sim p(\theta|Data)$, then use it to sample $\tilde{y}^{(1)} \sim p\left(\tilde{y} \mid \theta^{(1)}\right)$
  - From $\left\{ \left(\theta, \tilde{y}\right)^{(1)}, \ \ldots \left(\theta, \tilde{y}\right)^{(S)} \right\}$ use $\{ \tilde{y}^{(1)}, \ \ldots \tilde{y}^{(S)} \}$ as **marginal posterior distribution** of $\tilde{Y}$

Topic2. Predictive distributions and model checking

- Example: Poisson model for birthrate
  - age-40 woman without college degree child birth vs. with a degree

Paper and pencil (Analytic)

$$\Pr(\tilde{Y}_1 > \tilde{Y}_2 | \sum Y_{i,1} = 217, \sum Y_{i,2} = 66) =$$

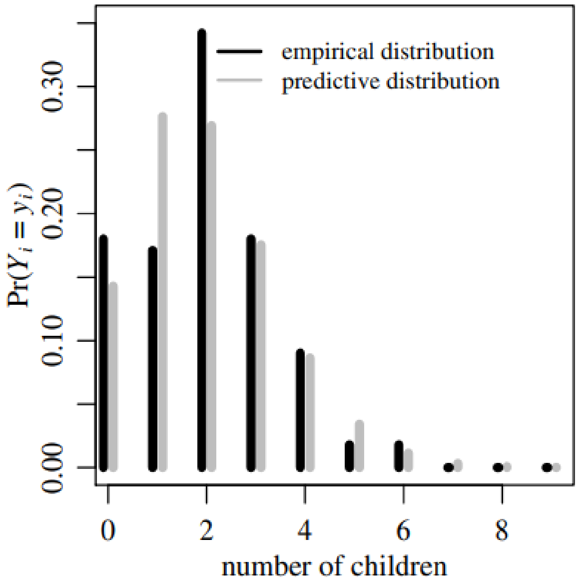
$$\sum_{\tilde{y}_2=0}^{\infty} \sum_{\tilde{y}_1=\tilde{y}_2+1}^{\infty} \text{dnbinom}(\tilde{y}_1, 219, 112) \times \text{dnbinom}(\tilde{y}_2, 68, 45)$$

Simulation (Monte Carlo)

```
> a<-2 ; b<-1
> sy1<-217 ;  n1<-111
> sy2<-66  ;  n2<-44

> theta1.mc<-rgamma(10000,a+sy1, b+n1)
> theta2.mc<-rgamma(10000,a+sy2, b+n2)
> y1.mc<-rpois(10000,theta1.mc)
> y2.mc<-rpois(10000,theta2.mc)

> mean(y1.mc>y2.mc)
[1] 0.4823
```

Topic2. Predictive distributions and model checking

- Model checking
  - 40 yr women w/o college degree, distn of # of children
  - Empirical: Out of 111 women, 38 women has exactly two children (twice women having one children)
  - Predictive: probability of sampling women with two child and one are about the same
  - Why?
    - Sampling variability?
    - Poisson model is wrong ?



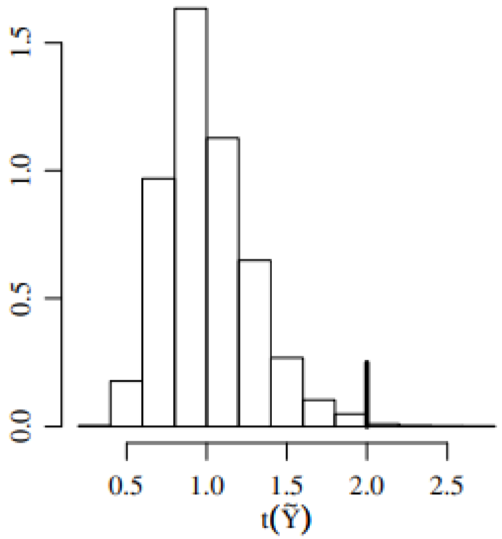
Topic2. Predictive distributions and model checking

- Model checking
  - $t(y)$: $\dfrac{\#(2\ child)}{\#\ (1\ child)}$

For each $s \in \{1, \ldots, S\}$,

$\{\theta^{(1)}, \ldots, \theta^{(S)}\}$ are samples from the posterior distribution of θ;
$\{\tilde{\mathbf{Y}}^{(1)}, \ldots, \tilde{\mathbf{Y}}^{(S)}\}$ are posterior predictive *datasets*, each of size n;
$\{t^{(1)}, \ldots, t^{(S)}\}$ are samples from the posterior predictive distribution of $t(\tilde{\mathbf{Y}})$.

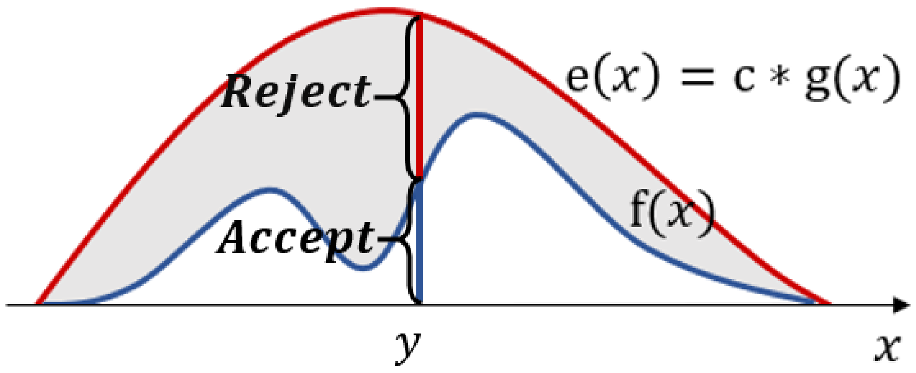
```
a<-2 ; b<-1
t.mc<-NULL

for(s in 1:10000) {
  theta1<-rgamma(1, a+sy1, b+n1)
  y1.mc<-rpois(n1, theta1)
  t.mc<-c(t.mc,sum(y1.mc==2)/sum(y1.mc==1))
                    }
```



Topic3. Rejection sampling, importance sampling

- Rejection Sampling
  - Target, f: difficult to generate but possible to evaluate
  - Proposal, g: Easy to generate
    - Envelope:$e(x) = constant\ * g(x)$

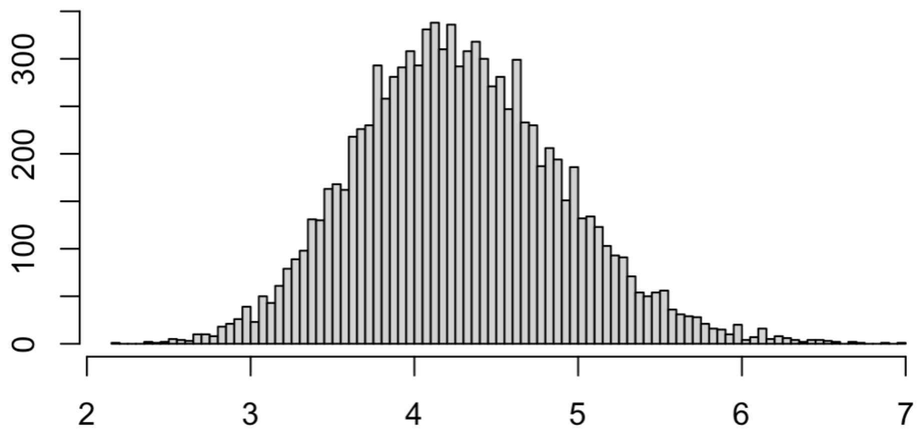
  - Algorithm:
    1. Sample $\mathrm{x} \sim \mathrm{g(x)}$
    2. Sample $\mathrm{U} \sim \mathrm{Uniform}\ (0,1)$
    3. Accept x if $\mathrm{U} < \frac{\mathrm{f(x)}}{\mathrm{e(x)}}$: acceptance ratio



Topic3. Rejection sampling, importance sampling

- Example: Sampling from posterior
  - Prior: $p(\theta) \sim \text{lognormal}\,(4, 0.5^2)$
  - Likleihood: $p(\text{Data} \mid \theta) \sim \text{Poisson}\,(\theta)$
  - Posterior: $p(\theta \mid \text{Data}) \propto p(\theta)p(\text{Data}|\theta)$
  - For envelope, use MLE $\bar{x}$ that maximizes likelihood
  - Acceptatnce ratio : $\frac{f}{e} = \frac{p(\theta)p(\text{data}|\theta)}{p(\theta)\,p(\text{data}|\text{MLE})} = \prod_i \frac{p(\text{data}_i|\theta)}{p(\text{data}|\text{MLE})}$

```r
x = c(8,3,4,3,1,7,2,6,2,7)
n = 10000
lambda.samp = rep(NA, n)
xbar = mean(x)
iter = 1; total = 1; # to check all iterations
while(iter <= n){
  lambda = exp(rnorm(1, log(4), 0.5))
  u = runif(1, 0, 1)
  ratio = exp(sum(dpois(x, lambda, log = TRUE)) - sum(dpois(x, xbar, log = TRUE)))
  if (u < ratio){
    lambda.samp[iter] = lambda
    iter = iter +1
  }
}
hist(lambda.samp, nclass = 100)
```



Topic3. Rejection sampling, importance sampling

- Importance Sampling
  - $p(\theta|y)$: hard to sample from -> consider unnormalized $\boldsymbol{q(\theta|y) = Const * p(\theta|y)}$
  - $g(\theta)$: easy to draw samples (envelope or proposal density)
    - $E[h(\theta)|y)] = \frac{\int h(\theta)p(\theta|y)d\theta}{\int p(\theta\ |y)\ d\theta} = \frac{\int h(\theta)q(\theta|y)d\theta}{\int q(\theta\ |y)d\theta} = \frac{\int \left[\frac{h(\theta)q(\theta)}{g(\theta)}\right]g(\theta)d\theta}{\int \frac{q(\theta)}{g(\theta)}g(\theta)d\theta}$
  - Algorithm:
    1. Sample $\theta^{(1)},\dots,\theta^{(S)}$ from $g(\theta)$
    2. Calculate importance ratio: $w\left(\theta^{(s)}\right) = \frac{q\left(\theta^{(s)}|y\right)}{g\left(\theta^{(s)}\right)}$ for each sample
    3. Calculate $\frac{\frac{1}{S}\sum_{s=1}^{S} h\left(\theta^{(s)}\right)w\left(\theta^{(s)}\right)}{\frac{1}{S}\sum_{s=1}^{S} w\left(\theta^{(s)}\right)}$
  - Note: q / g must be bounded and g have heavier tail than f
  - Importance ratio will appear later when calculating Metroplois Hastings ratio in MCMC
    - MH ratio $R\left(\theta^{(t)},\theta^*\right) = \frac{q(\theta^*)g\left(\theta^t\right)}{q(\ \theta^t\ )g(\theta^*)} = \frac{q(\theta^*\ )}{g(\theta^*)} * \frac{1}{\frac{q\left(\theta^t\ \right)}{g\left(\theta^t\right)}} = \frac{w(\theta^*)}{w(\theta^t)}$

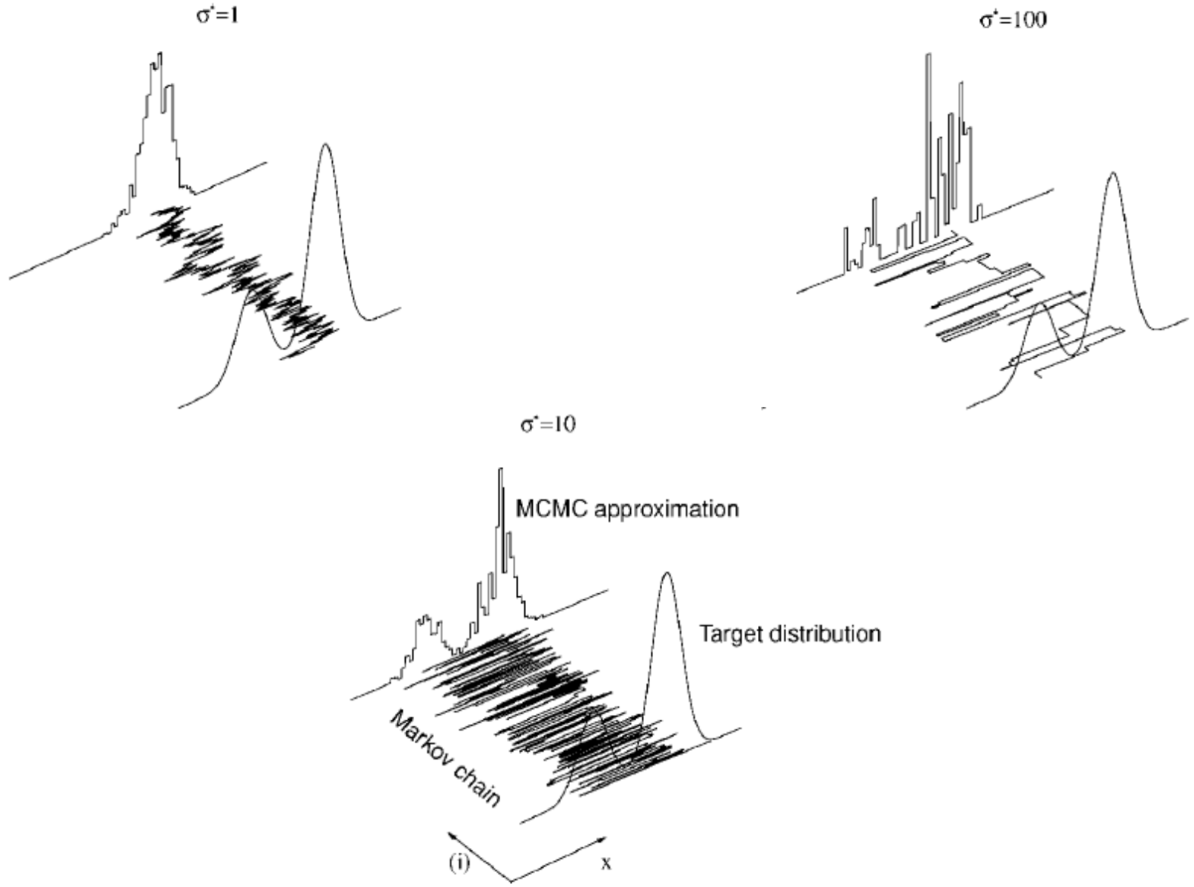
Why MCMC?

- Sometimes independent samples cannot be drawn easily…

- MC: independent sampling
- MCMC: dependent (smart) sampling

Some illustrations on MCMC

- Good proposal is important



HW

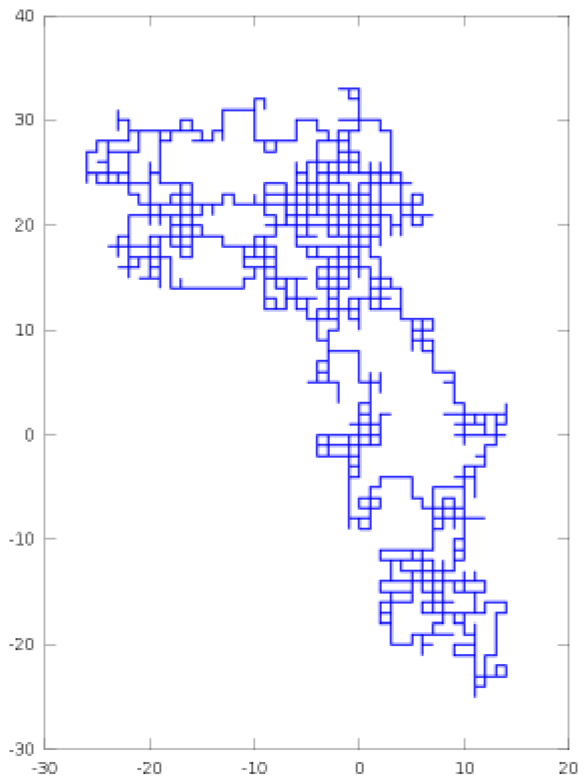
- FCB 4.7.

MCMC

ESC 6th week

최익준

Random walk



$$\eta_1, \eta_2 \ldots \sim i.i.d \;\; s.t \;\; \eta_i = \pm 1,$$

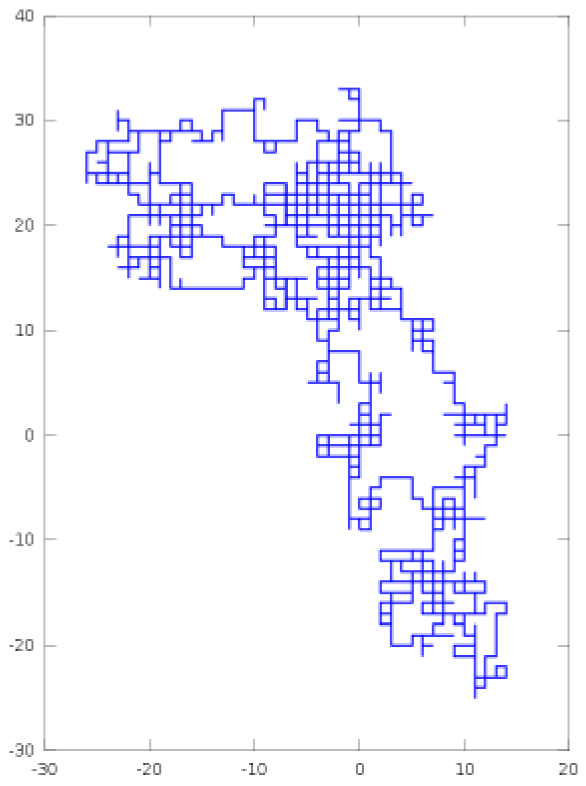
$$P(\eta_i = 1) = p, \qquad P(\eta_i = -1) = q$$

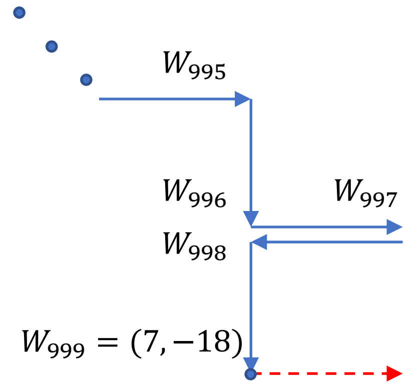
$$W_n = \eta_1 + \eta_2 + \cdots + \eta_n, W_0 := 0$$

$$P\big((a \pm 1, b) = W_{i+1} | (a, b) = W_i\big) = \frac{1}{4}$$

$$P\big((a, b \pm 1) = W_{i+1} | (a, b) = W_i\big) = \frac{1}{4}$$

Random walk

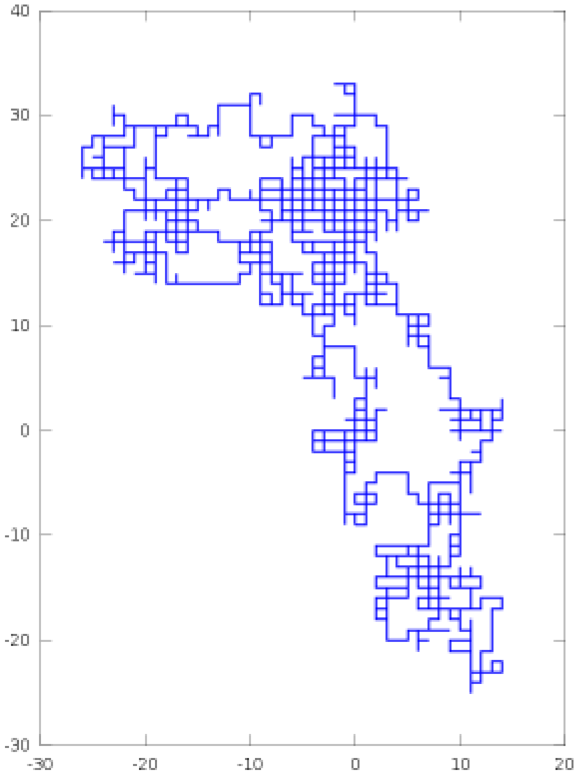


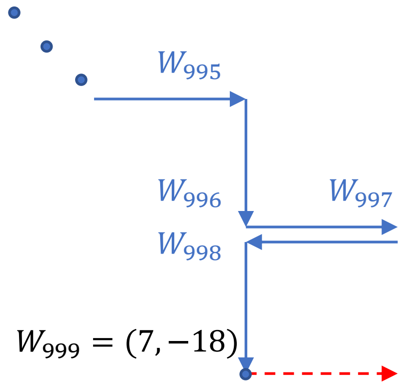


$$Stochastic\ process\ \{\ W_t\ \}_{t \in T}:$$
$$Collection\ of\ random\ variable\ with\ t, \omega$$

Discrete / *Continuous*

Random walk



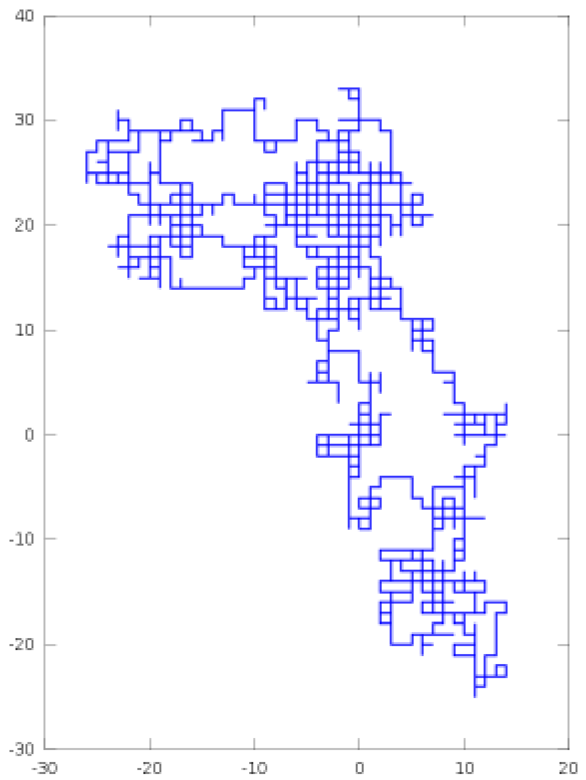


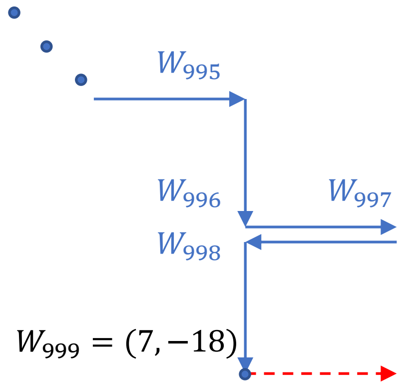
$$P(W_{1000} = (8, -18) \mid W_{999} = (7, -18), W_{998} = () \; \dots \; W_0) = \frac{1}{4}$$

$$\uparrow\downarrow$$

$$P(W_{1000} = (8, -18) \mid W_{999} = (7, -18)) = \frac{1}{4}$$

Markov chain





$$W_n \ depends \ only \ on \ W_{n-1}$$

$$Memoryless \ property:$$
$$P(\xi_{n+1} = s | \xi_0, \xi_1, \dots, \xi_{n-1}, \xi_i) = P(\xi_{n+1} = s | \xi_n)$$

Markov chain

$$\xi_n \; is \; called \; Markov \; Chain \; on \; S \; if \; \forall n, \forall s, :$$
$$P(\xi_{n+1} = s \big| \xi_0, \xi_1, \ldots, \xi_{n-1}, \xi_n) = P(\xi_{n+1} = s | \xi_n)$$

<u>Remind Poisson process (2nd. Week)</u>

Poisson process :$N^t(s) = N(t+s) - N(t) = N(s) \sim Poi(\lambda t) \;\; \forall s, t$

s : time increment

t : specific time

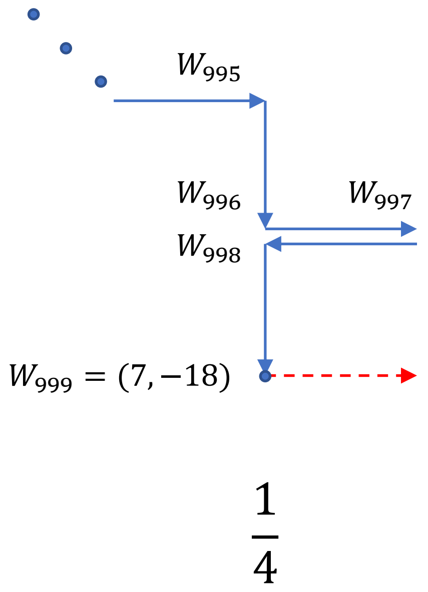
$$N(t+s) = N^t(s) + N(t)$$

Markov chain

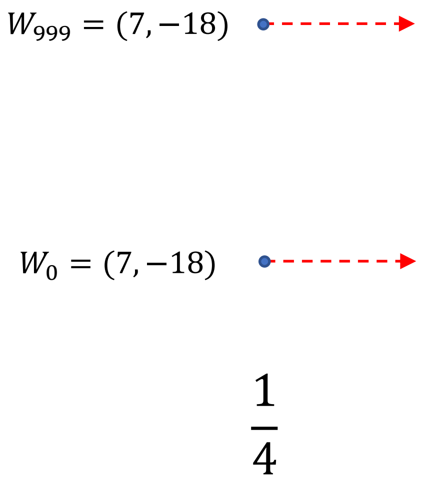
$$time\ homogeneous\ Markov\ chain: \forall n, \forall i, j\ ,$$

$$P(\xi_{n+1} = j | \xi_0, \xi_1, \dots, \xi_{n-1}, \xi_n = i) = \underbrace{P(\xi_{n+1} = j | \xi_n = i)}_{transition\ probability}$$

$$= \mathrm{P}(\xi_1 = j | \xi_0 = i)$$

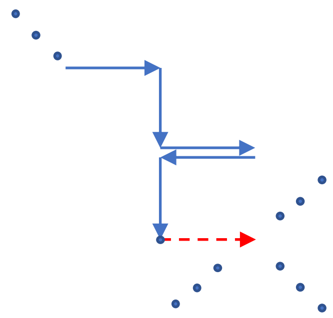


$\Leftrightarrow$



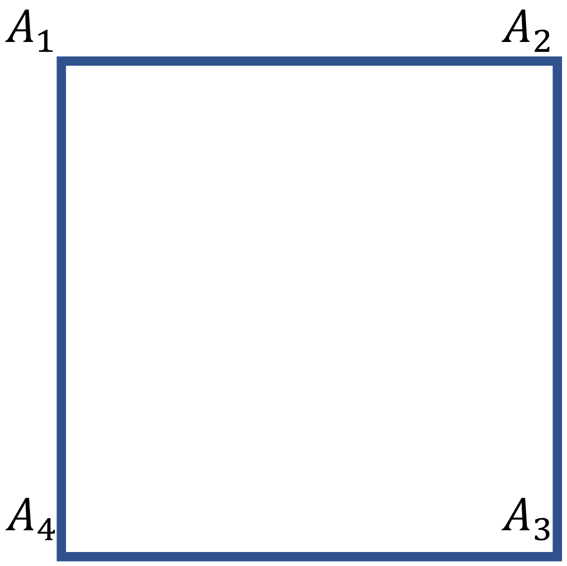
Markov chain

$$time\ homogeneous\ Markov\ chain: \forall n, \forall i, j\ ,$$
$$P(\xi_{n+1} = j | \xi_0, \xi_1, \dots, \xi_{n-1}, \xi_n = i) = P(\xi_{n+1} = j | \xi_n = i)$$
$$= \mathrm{P}(\xi_1 = j | \xi_0 = i)$$



$$Z \times Z = S \ni i, j\ (infinite)$$

Markov chain

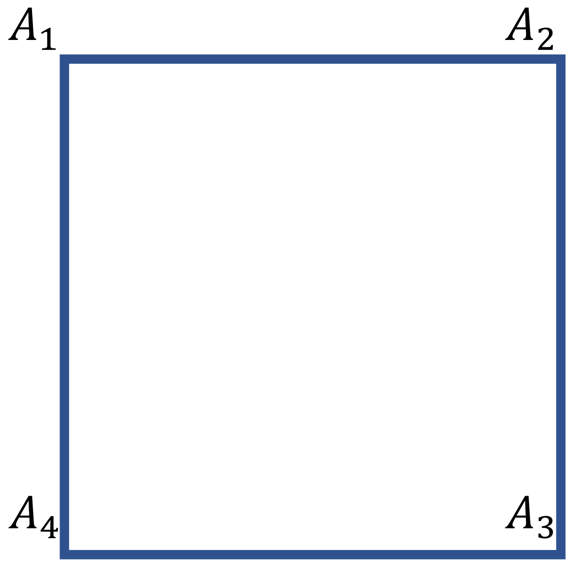


$S = \{A_1, A_2, A_3, A_4\}$ $(finite)$

$$P(clockwise\ 1step\ movement) = \frac{1}{2}$$

$$P(conter - clockwise\ 1step\ movement) = \frac{1}{2}$$

Markov chain



$$S = \{A_1, A_2, A_3, A_4\}\ (finite)$$

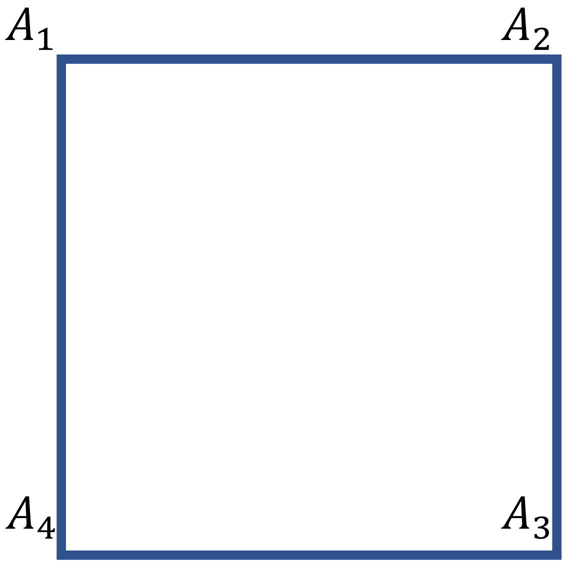
trasition prob. matrix

$$\begin{array}{c} \\ A_1 \\ A_2 \\ A_3 \\ A_4 \end{array}\begin{array}{c} \begin{array}{cccc} A_1 & A_2 & A_3 & A_4 \end{array} \\ \begin{bmatrix} 0 & 0.5 & 0 & 0.5 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.5 & 0 & 0.5 \\ 0.5 & 0 & 0.5 & 0 \end{bmatrix} \end{array} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{bmatrix}$$

transition probability

$$p_{ij} = P\left(W_{n+1} = A_j \,\middle|\, W_n = A_i\right)$$

Markov chain



$S = \{A_1, A_2, A_3, A_4\}\ (finite)$

$2 - step\ trans.prob.?$

$$P\left(W_{n+2} = A_j \,\middle|\, W_n = A_i\right)?$$

$ex)$
$P(W_{n+2} = A_1 \mid W_n = A_1)$
$= p_{11}p_{11} + p_{12}p_{21} + p_{13}p_{31} + p_{14}p_{41}$
$= 0 * 0 + 0.5 * 0.5 + 0 * 0 + 0.5 * 0.5 = 0.5$

$n - step\ trans.prob.?$

$ex)\ n = 3$
$P(W_{n+3} = A_1 \mid W_n = A_1) = \mathrm{p}_{11}\mathrm{p}_{11}\mathrm{p}_{11} + \mathrm{p}_{11}\mathrm{p}_{12}\mathrm{p}_{21} +$
$\dots + \mathrm{p}_{13}\mathrm{p}_{33}\mathrm{p}_{31}$

Markov chain

$2 - step\ trans.prob.?$

$$P(W_{n+2} = A_j \,|W_n = A_i)\,?$$

$ex)$
$P(W_{n+2} = A_4\,|W_n = A_2)$
$= p_{21}p_{14} + p_{22}p_{24} + p_{23}p_{34} + p_{24}p_{44}$
$= 0.5 * 0.5 + 0 * 0 + 0.5 * 0.5 + 0 * 0 = 0.5$

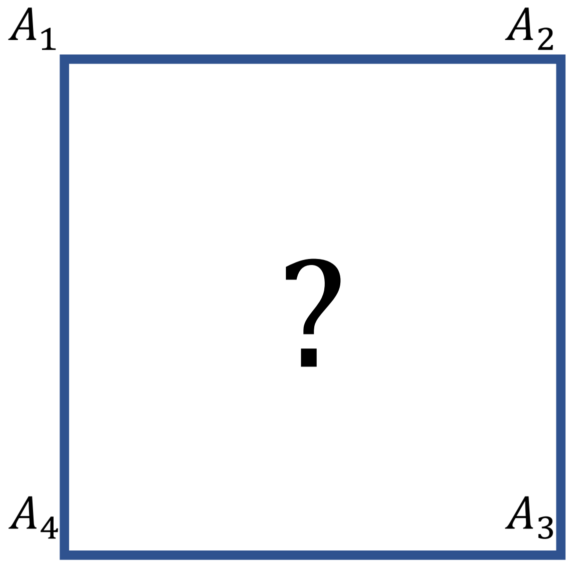
$$\begin{bmatrix} 0 & 0.5 & 0 & 0.5 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.5 & 0 & 0.5 \\ 0.5 & 0 & 0.5 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0.5 & 0 & 0.5 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.5 & 0 & 0.5 \\ 0.5 & 0 & 0.5 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{bmatrix} \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{bmatrix}$$

$\mathrm{n} - step\ trans.prob.?$

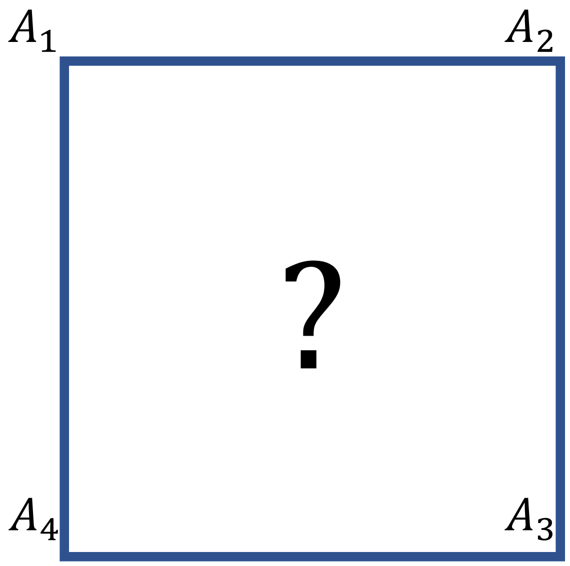
$$= \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{bmatrix} \cdots \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{bmatrix} \cdots \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{bmatrix} \Longrightarrow \text{Chapman} - \text{Kolmogorov eq.}$$

LTB of Markov chain



$after\ 100years\ \ldots$

LTB of Markov chain



$suppose\ W_0 = A_1$

$$P(W_1|W_0) = [0\ 0.5\ 0\ 0.5]$$
$$P(W_2|W_0) = [0.5\ 0\ 0.5\ 0]$$
$$\vdots$$

$odd\ or\ even$

$\Rightarrow$

$\frac{1}{4}$ $\frac{1}{4}$

$invariant\ measure$

$\frac{1}{4}$ $\frac{1}{4}$

LTB of Markov chain

Conditions for regularity…

- irreducible
- ergodic

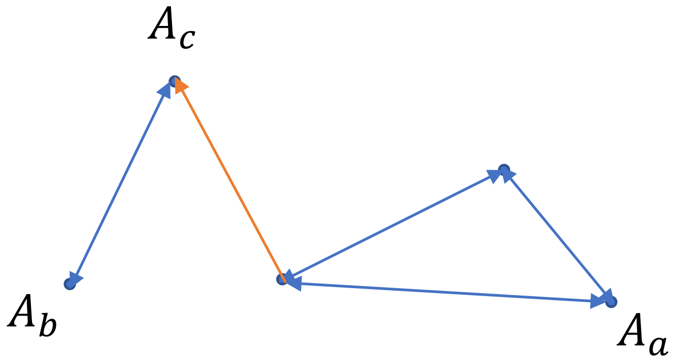
-positive recurrent

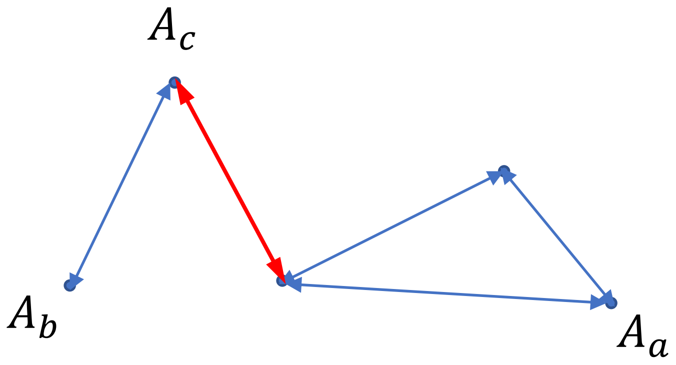
-aperiodic

=> Easy to investigate

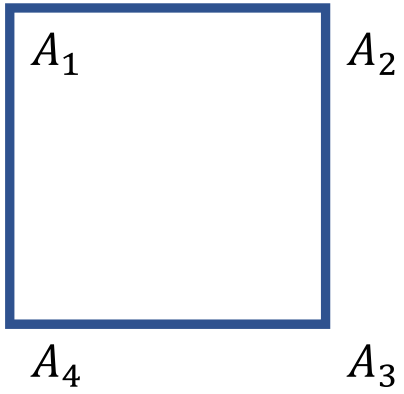
Long term behavior / limiting distribution

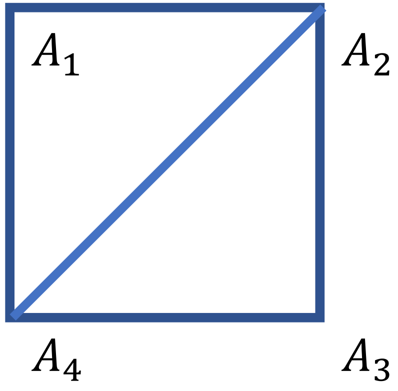
Regularity





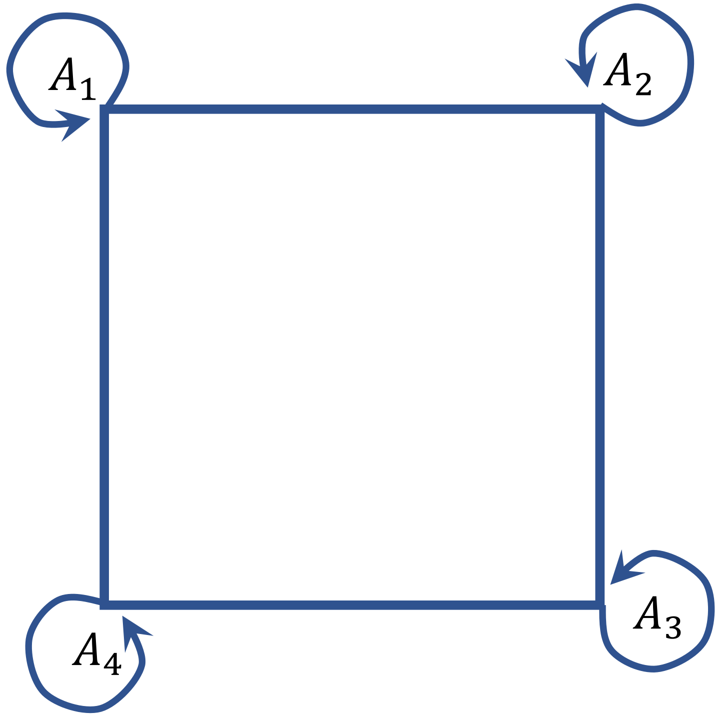
intercommunicate
irridicuble





aperiodic

Under regularity

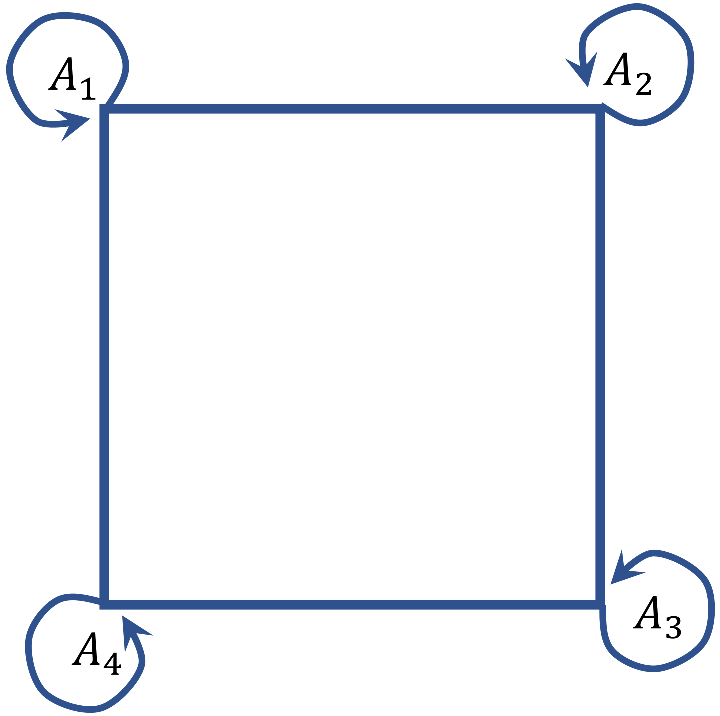


$$\begin{array}{c|cccc} & A_1 & A_2 & A_3 & A_4 \\ \hline A_1 & 0.2 & 0.4 & 0 & 0.4 \\ A_2 & 0.4 & 0.2 & 0.4 & 0 \\ A_3 & 0 & 0.4 & 0.2 & 0.5 \\ A_4 & 0.4 & 0 & 0.4 & 0.2 \end{array}$$

$100 years\ later \ldots$

$$\left.\begin{array}{l} P(W_{36500} = A_1) \\ P(W_{36500} = A_2) \\ P(W_{36500} = A_3) \\ P(W_{36500} = A_4) \end{array}\right. \neq 0$$

Under regularity



$$suppose\ W_0 = A_1$$

$\frac{1}{4}$ $\frac{1}{4}$

$$P(W_1|W_0) = [0.2\ 0.4\ 0\ 0.4]$$
$$P(W_2|W_0) = [0.36\ 0.16\ 0.32\ 0.16]$$
$$\vdots$$

$$\Longrightarrow \quad invariant\ measure\ \pi$$

$$\lim_{n\to\infty} P(W_n = A_i|W_0) = \pi_i$$

$$Consider\ each\ W_i : \ sampling\ !$$

$\frac{1}{4}$ $\frac{1}{4}$

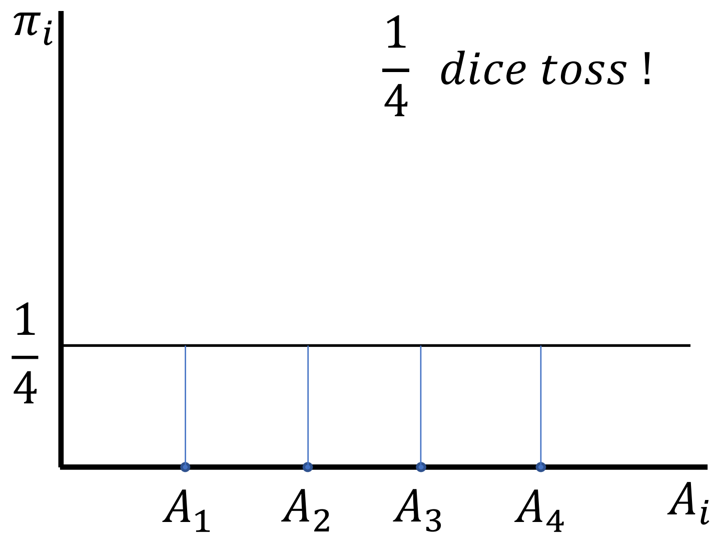
Limiting Distribution

$\frac{1}{4}$ $\frac{1}{4}$

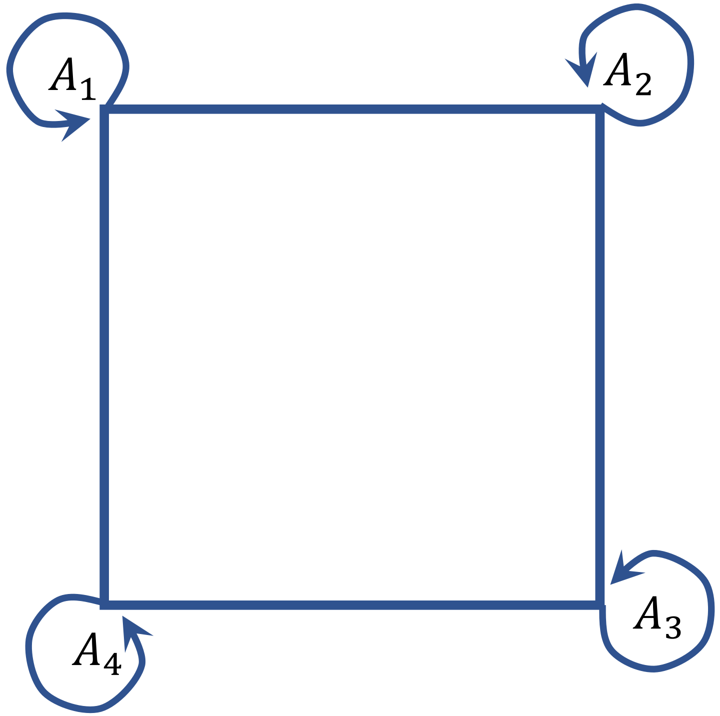
$invariant\ measure\ \pi_i$

$\frac{1}{4}$ $\frac{1}{4}$

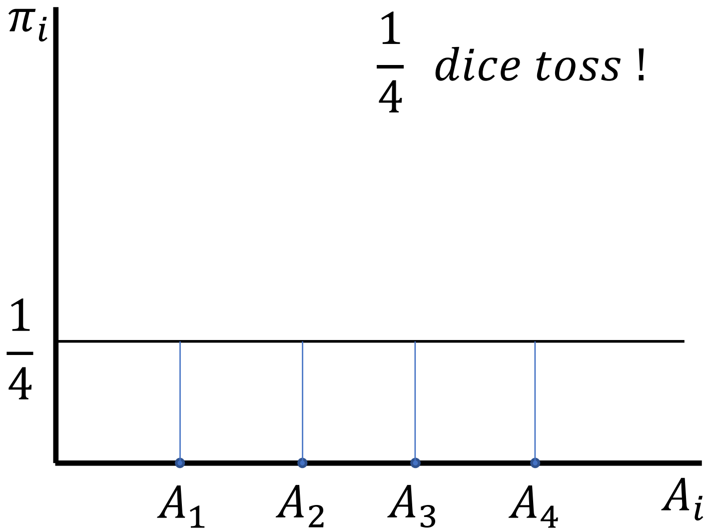
$$\xrightarrow{\sum \pi_i = 1}$$



MCMC



$p(i) = \pi_i$



MCMC

$$p(i) = \pi_i$$

ㅠㅠ $distribution \rightarrow proper\ markov\ chain$

$$\downarrow$$

$W_{i+1}\ after\ W_i$(sampling) $\rightarrow$ Approximation !

Metropolis − Hastings, HMC, … : How to build Markov chain which satisfies regularity conditions

Markov chain

Simulation

HW

Binomial $(N, \frac{1}{2})$?

HW

$Ehrenfest\ model$ (Dog – flea model)

-to explain the second law of thermodynamics..(Entropy)

Two dogs (Mac, Donald) share a population of N fleas. At each discrete unit of time, one of fleas jumps from the dog it is on to the other dog. Let X_n denote the number of fleas on Mac after n jumps. If there are i fleas on Mac, then on the next jump the number of fleas on Mac either goes up by one, if one of the $N - i$ fleas on Donald jumps to Mac, or goes down by one, if one of the i fleas on Mac jumps to Donald.

HW

At each unit time, one of the fleas jumps!

Q: # of fleas on Mac ?

Ex)

N=10

At $i-th$ period…

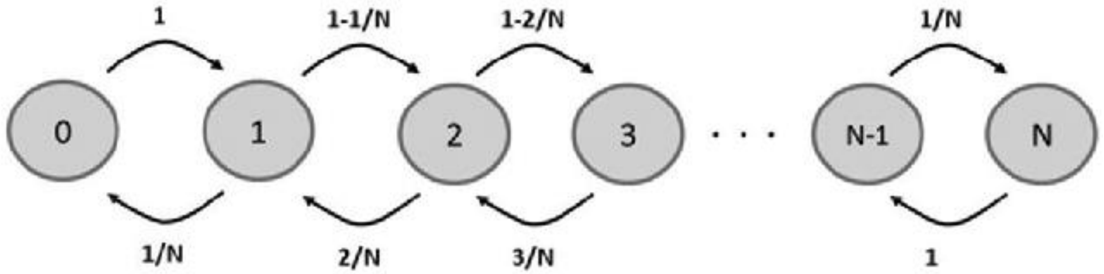
Mac 6 / 4 Donald

At $(i+1)-th$ period

Mac 7 / 3 Donald (with prob. = 4/10)

Mac 5 / 5 Donald (with prob. = 6/10)

$$p_{ij} = \begin{cases} \dfrac{i}{N} & if\ j = i-1 \\ \dfrac{N-i}{N} & if\ j = i+1 \\ 0 \quad . & otherwise \end{cases}$$



HW

Markov chain
-time homogeneous
-finite
Regular conditions
-irreducible
-ergodic

$$Modified\ Ehrenfest\ model$$
$$\#of\ total\ fleas = N$$
$$S = 0,1,2,\dots,N-1,N \ni i,j$$

$$p_{ij} = \begin{cases} \dfrac{i}{2N} & if\ j = i-1 \\ \dfrac{N-i}{2N} & if\ j = i+1 \\ \dfrac{1}{2}. & if\ j = i \\ 0 & .\quad o/w \end{cases}$$

HW

- Let N=5 on Modified Ehrenfest model (S={0,1,2,3,4,5})

(a) Find transition prob. matrix :=M

(b) Calculate $M^{10},\ M^{50},\ M^{100}$ and check each entry converges somewhere.

(c) Check ∀x ∈ S, $\mathrm{P}('\#\ of\ fleas\ on\ Mac\ at\ t=100'=x)$ is hardly affected by '# of fleas on Mac at t=0' (to say that $\mathrm{P}('\#\ of\ fleas\ on\ Mac\ at\ enough\ large\ t'=x)$ is not affected by '# of fleas on Mac at t=0' at all).

*Hint: initial state can be represented by row vector.

Ex) $(initial\ state: W_0=2)\ [0\ 1] * (2-step\ transition\ prob.matrix) \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}$

$= P(W_2=i|W_0=2)[p_{21}^{(2)}\ p_{22}^{(2)}]$

(d) Let T=50000, and consider each $W_{t\in T}$ as a sampling of distribution D. Check how D looks like and compare D with $X \sim Binomial(\ 5, ½\ )$

(e) Check $P('\#\ of\ fleas\ on\ Mac\ at\ t=100'=x) \cong P_D(x) \cong P_{Binomial\left(5,\frac{1}{2}\right)}(x)$

7th week session…

- Intercommunicate
- Recurrent, transient
- irreducible
- Periodic, aperiodic
- Detailed balance
- Ergodic
- Invariant measure / limiting distribution (eigenvector)