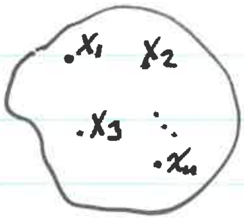
Chapter 2: Markov chain Monte Carlo

2.1 Introduction

MC



$p(X_1 X_2 \cdots X_n) = p(X_1)p(X_2)\cdots p(X_n)$

Independent samples

$S_L = \{X_i\}_{i=1}^{L}$

Histogram$(S_L) \approx p$

Sample p directly

MCMC

$p(X_1 X_2 \cdots X_n) = p(X_1)\, p(X_2|X_1) \cdots p(X_n|X_{n-1})$

Markov samples: $P(X \to X') = p(X'|X)$

$S_L = \{X_i\}_{i=1}^{L}$ 'trajectory'

Histogram$(S_L) \approx p$

Choose transition probability $P(X\to X')$ to sample p

2.2. Markov chains

Grimmett + Stirzaker Chapter

· Def: $(X_i)_{i=1}^{n}$ is a discrete-time Markov chain (MC) if

$$P(X_n = x_n \mid X_1 = x_1, X_2 = x_2, \dots, X_{n-1} = x_{n-1}) = \underbrace{P(X_n = x_n \mid X_{n-1} = x_{n-1})}_{\text{transition probability}}$$

for all $\underbrace{n \geq 1}_{\text{all sequences}}$ and all $\underbrace{x_1, x_2, \dots, x_n}_{\text{all values}}$.

$$\Rightarrow P(X_1, X_2, \dots, X_n) = P(X_1)P(X_2|X_1)\cdots P(X_n|X_{n-1})$$

$$X_1 \to X_2 \to \cdots \to X_n$$

· Assumptions / restrictions:

- $X_i \in \mathcal{X}$ discrete / countable state space
- $P(X_n = b \mid X_{n-1} = a)$ doesn't depend on $\underset{\text{time}}{n}$

$$\Rightarrow P(X_n = b \mid X_{n-1} = a) = P(X_2 = b \mid X_1 = a) \quad \forall n$$

Time-invariant, time-independent, homogeneous

· Transition matrix

$$P(X_n = j \mid X_{n-1} = i) = P(i \to j) = \Pi_{ij} \quad \text{or} \quad P_{ij}$$

- $|\mathcal{X}| \times |\mathcal{X}|$ matrix
- $0 \le \Pi_{ij} \le 1 \quad \forall\, i,j$
- $\sum_j \Pi_{ij} = 1 \quad \forall\, i$

} Stochastic matrix

$$\Pi = (\Pi_{ij}) = \begin{pmatrix} \Pi_{11} & \Pi_{12} & \Pi_{13} & \cdots \\ \Pi_{21} & \Pi_{22} & \Pi_{23} & \cdots \\ \vdots & & \ddots & \end{pmatrix}$$

row, column

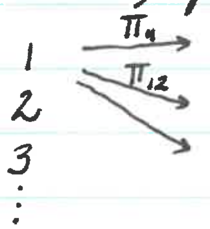
row sum = 1

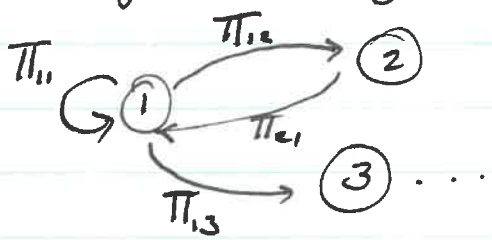
Rem: Different convention (used in physics)

$$P(i \to j) = \tilde{\Pi}_{ji} \qquad \sum_j \tilde{\Pi}_{ji} = 1 \quad \forall\, i$$

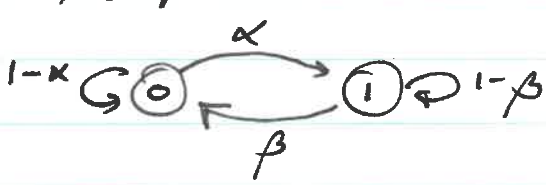
$$\tilde{\Pi} = \Pi^T$$

· Graphical representation: $\Pi_{ij} = P(i \to j)$



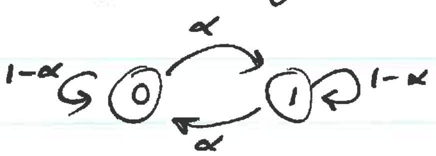


· Example: 2-state MC



$$\Pi = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix} \begin{matrix} \to \Sigma \cdots = 1 \\ \to \Sigma \cdots = 1 \end{matrix}$$

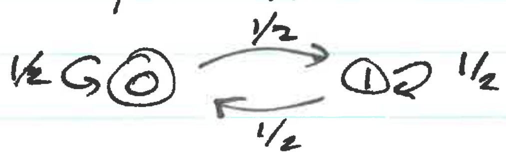
- $\beta = \alpha$: Symmetric MC



$$\Pi = \begin{pmatrix} 1-\alpha & \alpha \\ \alpha & 1-\alpha \end{pmatrix}$$

- $\alpha \approx 0$: 0 0 0 ⋯ 1 1 1 1 ⋯ 0 0 1 0 ⋯ — persistence: long sequences of 0's, 1's
- $\alpha \approx 1$: 0 1 0 1 ⋯ 0 0 1 1 0 1 ⋯ — anti persistence: alternating bits

- $\beta = \alpha = \frac{1}{2}$



$$\Pi = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$$

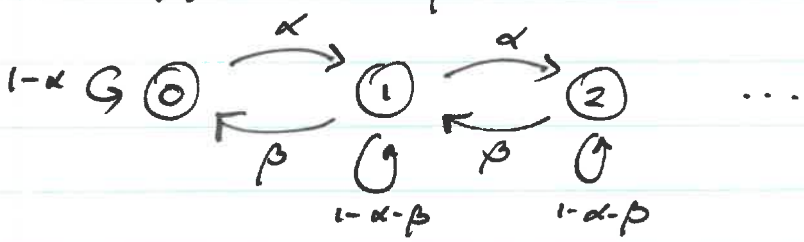
Coin tossing. No correlation

Rem: Independent MC if Π_{ij} doesn't depend on i:

$P(i \to j) = P(j)$ all rows the same

$\Rightarrow P(X_1 X_2 \ldots X_n) = P(X_1) P(X_2) \cdots P(X_n)$ iid model

- Example: Population model / birth-death process
  - $X_i \in \{0, 1, 2, \ldots\}$
  - $P(i \to i+1) = \alpha$ births
  - $P(i \to i-1) = \beta$ deaths



2.3 Probability propagation

$$X_1 \to X_2 \to \cdots \to X_n$$

$P(X_1)$ $P(X_n)$?

$$P(X_1, X_2) = P(X_1) P(X_2 | X_1)$$

$$\Rightarrow P(X_2 = j) = \sum_i P(X_1 = i, X_2 = j)$$
$$= \sum_i P(X_1 = i) P(X_2 = j | X_1 = i)$$

$$\Rightarrow P(X_{n+1} = j) = \sum_i P(X_n = i) P(X_{n+1} = j | X_n = i)$$
$$= \sum_i P(X_n = i) \Pi_{ij}$$

- Matrix notation:

$\bar{p}_n$ $\qquad (\bar{p}_n)_i = \bar{p}_n(i) = P(X_n = i)$ $\qquad \Sigma\, p_n(i) = 1$

Π $\qquad (\Pi)_{ij} = P(X_{n+1} = j \mid X_n = i)$ $\qquad \Sigma_j\, \Pi_{ij} = 1$

- Propagation formula:

$$\underbrace{\bar{p}_{n+1}}_{\text{row vector}} = \underbrace{\bar{p}_n}_{\text{row}} \; \underbrace{\Pi}_{\text{matrix}}$$

Chapman-Kolmogorov equation

$$\left(p_{n+1}(1) \quad p_{n+1}(2) \cdots \right) = \left(p_n(1) \quad p_n(2) \cdots \right) \left(\quad \Pi \quad \right)$$

$$\bar{p}_2 = \bar{p}_1 \Pi$$
$$\bar{p}_3 = \bar{p}_2 \Pi = \bar{p}_1 \Pi \Pi = \bar{p}_1 \Pi^2$$
$$\vdots$$
$$\bar{p}_n = \bar{p}_{n-1} \Pi = \bar{p}_1 \underbrace{\Pi^{n-1}}_{\text{n-1 step transition prob.}}$$

$$P(X_n = j) = \sum_i P(X_1 = i)\,(\Pi^{n-1})_{ij}$$
$$= \sum_i P(X_1 = i)\, \underbrace{P(X_n = j \mid X_1 = i)}_{\text{n-1 steps}}$$

$$\underset{P(X_1)}{X_1} \xrightarrow{\Pi} X_2 \xrightarrow{\Pi} \cdots \xrightarrow{\Pi} \underset{P(X_n)}{X_n}$$

- Example: 2-state MC

$$\begin{tikzcd} 0 \arrow[loop left, "1-\alpha"] \arrow[r, bend left, "\alpha"] & 1 \arrow[l, bend left, "\beta"] \arrow[loop right, "1-\beta"] \end{tikzcd} \qquad \Pi = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$$

$$\left(p_{n+1}(0) \quad p_{n+1}(1)\right) = \left(p_n(0) \quad p_n(1)\right) \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$$

$$\rightarrow \quad \begin{aligned} p_{n+1}(0) &= (1-\alpha)\, p_n(0) + \beta\, p_n(1) \\ p_{n+1}(1) &= \alpha\, p_n(0) + (1-\beta)\, p_n(1) \end{aligned}$$

Rem: Column vector convention

$$\begin{pmatrix} | \\ p_{n+1} \\ | \end{pmatrix} = \begin{pmatrix} \tilde{\Pi} \end{pmatrix} \begin{pmatrix} \bar{p}_n \end{pmatrix} \quad \text{vs} \quad \begin{pmatrix} p_{n+1} \end{pmatrix} = \begin{pmatrix} p_n \end{pmatrix} \begin{pmatrix} \Pi \end{pmatrix}$$

2.4. Ergodic Markov chains

- Stationary distribution:

$$p^* = p^* \Pi$$

  - Fixed point of Π: $p_1 = p^* \rightarrow p_2 = p_1 \Pi = p^* \Pi = p^*$
    $\Rightarrow p_n = p^* \;\; \forall n$
  - Eigenvector of Π with eigenvalue 1 (left eigenvect)
  - Π can have many stationary dist.

- Limiting distribution:

$$p_\infty = \lim_{n\to\infty} p_n = \lim_{n\to\infty} p_1 \Pi^{n-1}$$

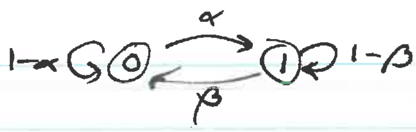
  - Might not exist
  - Might depend on p_1 choice of initial condition

- Ergodic Markov chain:
  - p_∞ exists
  - p_∞ independent of p_1
  - $p_\infty = p^*$

Grimmett + Stirzaker Secs 6.3 6.4

- Proposition: If $(X_i)_{i=1}^n$ is an aperiodic and irreducible Markov chain, then $\lim_{n\to\infty} p_n = p^*$ for any p_1.

· Example: 2-state MC



· $0 < \alpha, \beta < 1$: Ergodic $\bar{p}^* = \left(\frac{\beta}{\alpha+\beta} \quad \frac{\alpha}{\alpha+\beta} \right)$

· $\alpha = \beta = 0$: Not ergodic

Any p_1 is stationary

Reducible Markov chain

· $\alpha = \beta = 1$: Not ergodic

Periodic Markov chain

$p_1 = (a \quad b)$

$p_2 = (b \quad a)$

$p_3 = (a \quad b)$

$\vdots$

$p = \left(\frac{1}{2} \quad \frac{1}{2} \right)$ stationary but not a limiting distribution

· Ergodic theorem: If $(X_i)_{i=1}^n$ is an ergodic Markov chain, then

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} g(X_i) = E_{p^*}[g(X)] \quad \text{in probability}$$

estimator / time average: $\hat{\gamma}_n$ — stationary expectation: γ

$$\lim_{n\to\infty} P(|\hat{\gamma}_n - \gamma| > \varepsilon) = 0$$

· Generalization of law of large numbers (LLN) to Markov chains.

$X_i \sim p^*$ iid	$(X_i)_{i=1}^n$ MC
$\frac{1}{n}\sum_{i=1}^{n} g(X_i) \to E_{p^*}[g(X)]$ in prob.	$\frac{1}{n}\sum_{i=1}^{n} g(X_i) \to E_{p^*}[g(X)]$ in prob.

· Empirical occupation:

$$\hat{p}_n(j) = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i,j} = \text{fraction of time spent in } j$$

$$\delta_{i,j} = \begin{cases} 1 & i=j \\ 0 & i\neq j \end{cases}$$

$$\hat{p}_n(j) \xrightarrow{n\to\infty} p^*(j) \quad \text{in prob.}$$

· Interpretations of p^*:

1- Distribution of X_n as $n\to\infty$

2- Long time (stationary) occupation

2.5. Stationary distribution

$$p^* = p^* \Pi$$

1- Direct calculation

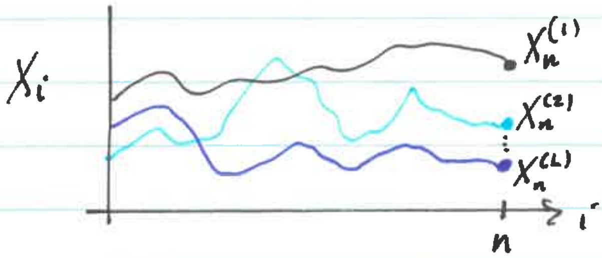
- Input: Π
- Output: left eigenvector of Π with eigenvalue 1
- Normalization: $\sum_i p^*(i) = 1$

· Code:

```
import scipy.linalg as la
pimat = np.array([[a,b,c], [...], [...]])
eigvals, eigvecs = la.eig(np.transpose(pimat))
```

≠ np why?

2- Parallel simulations



- Simulate L copies/realizations/trajectories

$$\left\{ \left(X_i^{(j)}\right)_{i=1}^{n} \right\}_{j=1}^{L}$$

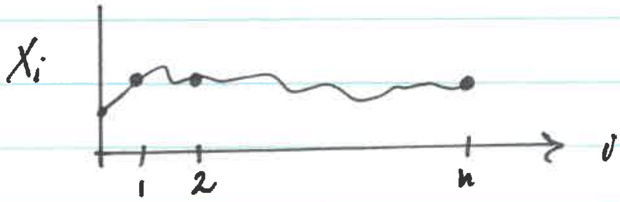
L = sample size / no traj.
n = final time

- Keep last state at time n: $\left\{ X_n^{(j)} \right\}_{j=1}^{L}$
- Histogram: $\hat{p}_{n,L}(i) \approx p_n(i)$ All final states (LLN)
- Convergence: $\hat{p}_{n,L}(i) \xrightarrow[L\to\infty]{n\to\infty} p^*(i)$

· Code:

```
n = 100
L = 10**5
xsample = []
for j = 1:L
    x = initial value        ← deterministic or random
    for i = 1:n
        generate new x                  2 loops
    append x to xsample
histogram(xsample)                      n → ∞, L → ∞
```

3- Ergodic simulation



- Simulate 1 trajectory $(X_i)_{i=1}^{n}$
- Empirical occupation:

$$\hat{p}_n(j) = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i,j}$$ Histogram of all state in time

- Convergence: $\hat{p}_n(j) \xrightarrow{n\to\infty} p^*(j)$ Ergodic theorem

· Code:

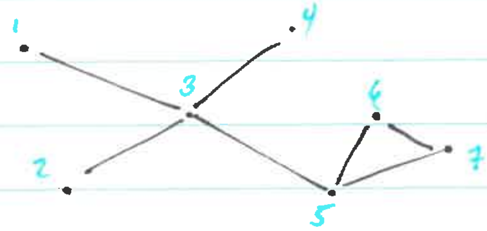
```
n = 10**3
xsample = []
x = initial value        deterministic or random
for i = 1:n
    generate new x                  1 loop!
    append x to xsample

histogram(xsample)
```

$n \to \infty$

Trade off n vs $n \times L$

2.6. Application: Random walk on graphs



- Graph: $G = (V, E)$ — vertices / nodes, edges
- Undirected
- Connected

· Adjacency matrix: $A_{ij} = \begin{cases} 1 & i \sim j \\ 0 & i \not\sim j \end{cases}$ connected / unconnected

- $|V| \times |V|$ matrix
- Symmetric $A = A^T$ $i \sim j \Rightarrow j \sim i$ reflexive

· Node degree: $k_i = \#$ links to node i
$= \sum_j A_{ij}$

· Degree list: $\vec{k} = (k_1 \; k_2 \; k_3 \dots k_{|V|})$

· Number of edges: $M = \frac{1}{2} \sum_i k_i = \frac{1}{2} \sum_{ij} A_{ij}$

· Uniform random walk (URW):
1- Start at node $X_1 = i$
2- Choose node connected to i, random uniform
k_i of them $\Rightarrow P(i \to j) = \frac{1}{k_i}$ $j \sim i$
3- Repeat

· Transition matrix: $\Pi_{ij} = P(i \to j) = \frac{A_{ij}}{k_i}$

- $0 \leq \Pi_{ij} \leq 1 \quad \forall i, j$
- $\sum_j \Pi_{ij} = \frac{1}{k_i} \sum_j A_{ij} = 1 \quad \forall i$
- Π ergodic if G connected

· Stationary / ergodic distribution: $p^*(i) = \frac{k_i}{2M} \propto k_i$
· Normalization: $\sum_i p^*(i) = \frac{1}{2M} \sum_i k_i = 1$

- Occupation $\hat{p}_n(i) \longrightarrow p^*(i) \propto k_i$
- Basis of Google Page rank

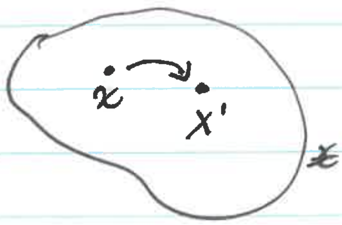
2.7. Metropolis-Hastings algorithm

Liu Chap. 5

- Refs: · Metropolis, Rosenbluth^2, Teller^2, 1953 (1948) $M(RT)^2$
  · Hastings 1970
- Top 10 algo of 20th century by SIAM

- Goal: · Generate variates/values $X \sim f$ target distribution
  · $x \in \mathcal{X}$ state space known, calculatable
  · $\sum_x f(x) = 1$

- Idea: Simulate ergodic Markov chain that has f as its stationary distribution.



- Steps:
  1- Initialize: $X_1 = x$ deterministic or random
  2- Generate move/proposal:
  $x \to x'$ with probability $q(x'|x)$
  $x' \sim q(\cdot|x)$
  3- Accept proposal with probability

$$\rho = \min\left\{1, \frac{q(x|x')f(x')}{q(x'|x)f(x)}\right\}$$

ρ_{MH} Metropolis-Hastings ratio

That is, $u \sim \mathcal{U}[0,1]$
if $u < \rho$ then
$X_2 = x'$ accept
else
$X_2 = x$ reject

4- Repeat

- Metropolis-Hastings MC:
  - $\pi(x'|x) = P(x \to x') = \begin{cases} \rho & x \neq x' \\ 1-\rho & x' = x \end{cases}$

  - Ergodic MC
  - Stationary distribution: $f(x)$ See CW2

- Remarks:
  - Free to choose $q(x'|x)$ Possible moves
  - Propose moves x' such that $f(x')>0$
  - $\mathrm{supp}(f) \subseteq \bigcup_x \mathrm{supp}(q(\cdot|x))$

  - Move accepted for sure if $q(x|x')f(x') > q(x'|x)f(x)$
    Move accepted with probability ρ otherwise

  - Implementation: if $\rho_{MH} \geq 1$ or $u < \rho_{MH}$ then accept
    check first — saves possibly generation of u

  - Or: if $\log \rho_{MH} > 0$ or $\log u < \log \rho_{MH}$

  - $f(x)$ needed only up to normalization factor: $f(x) = c\,g(x)$
    $$\Rightarrow \rho_{MH} = \frac{q(x|x')\,g(x')}{q(x'|x)\,g(x)}$$

  - Choose q such that accepted fraction $\approx \frac{1}{2}$
    acceptance ratio

· Particular cases:

1- Independent random walk: $q(x'|x) = q(x')$

$$\Rightarrow \rho_{MH} = \frac{q(x)\, f(x')}{q(x')\, f(x)}$$

2- Metropolis algorithm: $q(x'|x) = q(x|x')$ symmetric

$$\Rightarrow \rho_M = \frac{f(x')}{f(x)} \qquad \rho = \min\left\{1, \frac{f(x')}{f(x)}\right\}$$

· Accept move x' for sure if $f(x') \geq f(x)$
· Accept with prob. ρ if $f(x') < f(x)$

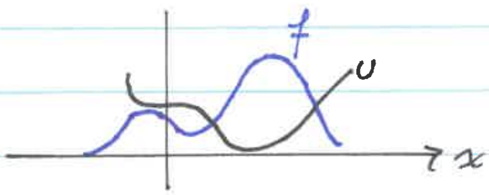
· Reversible: $\underbrace{f(x)}_{\text{high}}\, \underbrace{P(x \to x')}_{\text{low}} = \underbrace{f(x')}_{\text{low}}\, \underbrace{P(x' \to x)}_{\text{high}}$ CW2

3- Metropolis random walk:
· Move: $x' = x + \delta x$
· Displacement: $\delta x \sim p$ symmetric

$\Rightarrow q(x'|x) = q(x|x')$

$q(x'|x) = q(x \to x') = p(\delta x)$
$q(x|x') = q(x' \to x) = p(-\delta x)$

4- Heat bath algorithm:
· Target: $f(x) = \dfrac{e^{-\beta U(x)}}{Z}$



· $x \in \mathbb{R}^d$ or $\mathcal{X}$
· $U : \mathcal{X} \to \mathbb{R}$ potential
· $\beta \in \mathbb{R}$ usually $\beta > 0$ inverse temperature

$$\Rightarrow \rho_M = \frac{e^{-\beta U(x')}}{e^{-\beta U(x)}} = e^{-\beta \Delta U}$$

$\Delta U = U(x') - U(x)$ potential change

⇒ · Move accepted for sure if $\Delta U \leq 0$: $U(x') \leq U(x)$
· " " with prob. p_M if $\Delta U > 0$: $U(x') > U(x)$
· Potential minimization (not greedy)
· Sample where $f(x)$ is large ⇔ $U(x)$ is low

· Example : $f \sim \mathcal{N}(\mu, \sigma^2)$

```
n = 10**3                              MC steps
xsample = [ ]
x = initial value                      deterministic or random
for i=1:n
    disp ~ U[-a,a]                     symmetric, a = size
    xp = x + disp
    pM = f(xp) / f(x)
    if pM>1 or rand()<pM then
        x = xp                         accept
                                       ← no else
    append x to xsample                same x added/kept
                                       if xp not accepted
```

See demo

· MCMC estimation : $E_f[g(X)] = \gamma$
  · Trajectory : $(X_i)_{i=1}^{n}$
  · Estimator : $\hat{\gamma}_n = \frac{1}{n} \sum_{i=1}^{n} g(X_i)$
  · Ergodic thm : $\hat{\gamma}_n \to \gamma$ in prob. as $n \to \infty$

· Comparison :

Standard MC	MCMC
· iid variates/samples	· MC variates
· No correlation	· Correlated samples
· Difficult in high dim	· Efficient in high dim
· Clever method needed	· Generic / "black box"

Bailer-Jones Chaps 8, 9

2.8 MCMC inference

- Probabilistic model : $P(x \mid \theta)$ $x \in \mathcal{X}$ state, $\theta \in \Omega$ parameters

- Parameter posterior :

$$\underset{\text{data}}{P(\theta \mid D)} = \frac{\overset{\text{likelihood}}{P(D \mid \theta)} \; \overset{\text{prior}}{P(\theta)}}{\underset{\text{evidence}}{P(D)}}$$

- Likelihood : $P(D \mid \theta) = P(x_1 x_2 \ldots x_d \mid \theta)$
  $= \prod_i P(x_i \mid \theta)$ Naive Bayes

- Evidence :

$$P(D) = \int d\theta \; P(D \mid \theta) P(\theta)$$

- Tasks:

1- Parameter inference: See CW2
- $P(\theta \mid D)$ not known in closed form
- Sample from $P(\theta \mid D)$
- Propose move $\theta \to \theta'$
- Accept with prob. $\rho = \min\left\{1, \frac{P(\theta' \mid D)}{P(\theta \mid D)}\right\}$ Under Metropolis
- Data given

$$= \min\left\{1, \frac{P(D \mid \theta') P(\theta')}{P(D \mid \theta) P(\theta)}\right\}$$ No $P(D)$

2- Data generation
- Find best θ^* from D (e.g. most probable)
- Sample from $P(D \mid \theta^*)$
- Propose move $D \to D'$
- Used to complete missing data

3- Evidence calculation / model comparison
- Estimate integral $P(D) = \int d\theta \; P(D \mid \theta) P(\theta)$
- MC or MCMC

2.9. Error analysis

- Expectation: $\gamma = E_f[g(X)]$ $\quad X \sim f$ target distribution
- Estimator: $\hat{\gamma}_n = \frac{1}{n} \sum_{i=1}^{n} g(X_i)$ $\quad (X_i)_{i=1}^{n}$ ergodic MC
- Error:

$$\hat{\gamma}_n \pm \sigma_n \qquad \text{ci } 68\%$$
$$\hat{\gamma}_n \pm 2\sigma_n \qquad \text{ci } 95\%$$

$$\sigma_n = \sigma(\hat{\gamma}_n) = \sqrt{\mathrm{var}(\hat{\gamma}_n)}$$

$$\mathrm{var}(\hat{\gamma}_n) \sim \frac{1}{n} \quad \text{but} \neq \quad \underbrace{\frac{\sigma^2(g(X))}{n}}_{\text{iid variance/naive variance}}$$

- Methods:

1- Independent simulations / runs

- Simulate L independent MCs: $\left\{ (X_i^{(j)})_{i=1}^{n} \right\}_{j=1}^{L}$
- Sample of estimators: $\left\{ \hat{\gamma}_n^{(j)} \right\}_{j=1}^{L} = S$
- Mean estimator: $\hat{\hat{\gamma}}_{n,L} = \frac{1}{L} \sum_{j=1}^{L} \hat{\gamma}_n^{(j)}$
- Error: $\sigma_n = \sqrt{\mathrm{var}(S)}$
- Variance: $\mathrm{var}(S) = \underbrace{\frac{1}{L-1} \sum_{j=1}^{L} \left(\hat{\gamma}_n^{(j)} - \hat{\hat{\gamma}}_{n,L} \right)^2}_{\text{Standard iid variance}}$

2- Ergodic simulation

- Simulate 1 trajectory: $(X_i)_{i=1}^{n}$
- Estimator: $\hat{\gamma}_n = \frac{1}{n} \sum_{i=1}^{n} g(X_i)$
- $\mathrm{var}(\hat{\gamma}_n)$?

$$\mathrm{var}(\hat{\gamma}_n) = \mathrm{var}\left(\frac{1}{n}\sum_{i=1}^{n} g(X_i)\right)$$
$$= \frac{1}{n^2}\mathrm{var}\left(\sum_{i=1}^{n} g(X_i)\right) \quad \neq \sum_i \mathrm{var}(g(X_i))$$
$$= \frac{1}{n^2}\sum_{i,j=1}^{n} \mathrm{cov}(g(X_i), g(X_j))$$

$$\mathrm{cov}(X,Y) = E[(X-E[X])(Y-E[Y])] \quad \text{Covariance}$$
$$\mathrm{cov}(X,X) = \mathrm{var}(X) \quad \text{Variance}$$

$$\mathrm{var}(\hat{\gamma}_n) = \frac{1}{n^2}\left(\sum_{i=1}^{n}\mathrm{var}(g(X_i)) + \sum_{i\neq j}\mathrm{cov}(g(X_i), g(X_j))\right)$$
$$= \frac{1}{n^2}\left(\sum_{i=1}^{n}\mathrm{var}(g(X_i)) + 2\sum_{1\leq j<k\leq n}\mathrm{cov}(g(X_j), g(X_k))\right)$$

- $(X_i)_{i=1}^{n}$ ergodic so $P_n(j) = P(X_n = j) \to f(j)$ as $n \to \infty$.

$$\Rightarrow \mathrm{var}(\hat{\gamma}_n) \sim \frac{1}{n^2}\Big[n\,\underbrace{\mathrm{var}_f(g(X))}_{\text{naive variance}} + 2n\sum_{k=1}^{\infty}\underbrace{\mathrm{cov}_f}_{\text{MC started at } f}(g(X_1), g(X_{1+k}))\Big]$$

$$= \frac{\mathrm{var}_f(g(X))}{n}\Big[\underbrace{1 + 2\sum_{k=1}^{\infty}\mathrm{corr}(g(X_1), g(X_{1+k}))}_{\tau(g)}\Big]$$

$$\mathrm{corr}(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sigma_X\,\sigma_Y} \quad \text{auto correlation}$$

$$\Rightarrow \mathrm{var}(\hat{\gamma}_n) = \frac{\mathrm{var}_f(g(X))}{n}\,\tau(g)$$
$$= \frac{\sigma_f^2(g(X))}{n_{eff}}$$

$\sigma_f^2(g(X))$ ~ naive variance

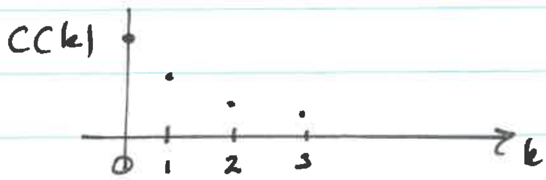
n_{eff} ~ effective number of samples

$$n_{eff} = \frac{n}{\tau(g)}$$

$\tau(g)$ ~ time scale of exponential decay of correlation

· Example: $\mathrm{corr}(g(X_1), g(X_{1+k})) = c(k) = \rho^k$, $\rho \in (0,1)$

$$\Rightarrow \quad 1 + 2\sum_{k=1}^{\infty} c(k) = \frac{1+\rho}{1-\rho}$$



3- Batch means

- n samples
- L batches/blocks of m points $\qquad L \times m = n$
- Block average: $\hat{\gamma}_j^m = \frac{1}{m} \sum_{j^{th} \text{ block}} g(X_k)$
- Estimator: $\hat{\gamma}_n = \frac{1}{L} \sum_{j=1}^{L} \hat{\gamma}_j^m$
- $n \to \infty$, $L \to \infty$ while $m \to \infty$, blocks become independent

$$\Rightarrow \mathrm{var}(\hat{\gamma}_n) = \text{naive var}(\{\hat{\gamma}_j^m\})$$