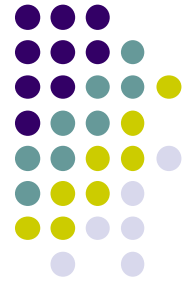
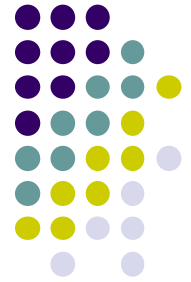


# Outline



- Motivation
  - Sampling from non-standard distribution
- Metropolis-Hastings Algorithm
  - Key idea
  - Implementation
- Example
  - Nonlinear SEM

# Basic Sampling Techniques



- Inverse method
- Rejection method
- Importance sampling
- Gibbs sampling
- . . . . .

# Motivated Example



$$\mathbf{y}_i = \mathbf{A}\mathbf{c}_i + \mathbf{\Lambda}\boldsymbol{\omega}_i + \boldsymbol{\epsilon}_i,$$

$$\boldsymbol{\eta}_i = \mathbf{B}\mathbf{d}_i + \mathbf{\Pi}\boldsymbol{\eta}_i + \mathbf{\Gamma}\mathbf{F}(\boldsymbol{\xi}_i) + \boldsymbol{\delta}_i,$$

where  $\mathbf{F}(\boldsymbol{\xi}_i)$  is a vector-valued nonlinear function of  $\boldsymbol{\xi}_i$ , and the definitions of other random vectors and parameter matrices are the same as before.



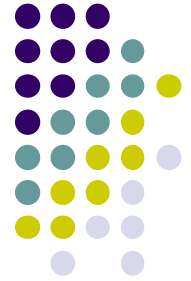
# Full Conditional Distribution

$$p(\Omega|\mathbf{Y}, \theta) = \prod_{i=1}^n p(\omega_i|\mathbf{y}_i, \theta) \propto \prod_{i=1}^n p(\mathbf{y}_i|\omega_i, \theta)p(\eta_i|\xi_i, \theta)p(\xi_i|\theta).$$

As  $\omega_i$  are mutually independent, and  $\mathbf{y}_i$  are also mutually independent given  $\omega_i$ ,  $p(\omega_i|\mathbf{y}_i, \theta)$  is proportional to

$$\exp \left\{ -\frac{1}{2}\xi_i^T \Phi^{-1} \xi_i - \frac{1}{2}(\mathbf{y}_i - \mathbf{A}\mathbf{c}_i - \Lambda\omega_i)^T \Psi_\epsilon^{-1}(\mathbf{y}_i - \mathbf{A}\mathbf{c}_i - \Lambda\omega_i) \right. \\ \left. - \frac{1}{2}(\eta_i - \mathbf{B}\mathbf{d}_i - \Pi\eta_i - \Gamma\mathbf{F}(\xi_i))^T \Psi_\delta^{-1}(\eta_i - \mathbf{B}\mathbf{d}_i - \Pi\eta_i - \Gamma\mathbf{F}(\xi_i)) \right\}.$$

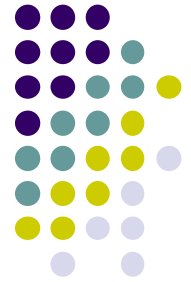
**Non-standard and complex!**



# Seminal Works

- [Metropolis, N. et al.](#) Equations of state calculations by fast computing machine. *Journal of Chemical Physics*, 1953, **21**, 1087–1091.
- [Hastings, W. K.](#) Monte Carlo sampling methods using Markov chains and their application. *Biometrika*, 1970, **57**, 97–109.

# Metropolis-Hastings Algorithm



- Objective

- Sample from a known (but nonstandard) density function:

$$X \sim f(x)$$

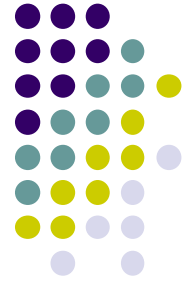
- Key idea

- Sample a candidate from a standard distribution (proposal distribution)
- Accept the candidate with some probability (acceptance probability)



# Implementation

1. Give an initial value:  $X^{(0)}$ .
2. At the  $l^{\text{th}}$  MH iteration with a current value  $X^{(l)}$ , a new candidate  $X_{\text{can}}$  is sampled from a proposal distribution  $q(\cdot | X^{(l)})$  which depends on  $X^{(l)}$ .



# Implementation

3. Calculate the acceptance probability

$$p_{accept} = \min \left\{ 1, \frac{f(X_{can})q(X^{(l)}|X_{can})}{f(X^{(l)})q(X_{can}|X^{(l)})} \right\}.$$

4. Generate  $u \sim U(0,1)$ ;

$$X^{(l+1)} = \begin{cases} X_{can}, & \text{if } u \leq p_{accept}, \\ X^{(l)}, & \text{otherwise.} \end{cases}$$

5. Repeat Steps 2.~4. until convergence.

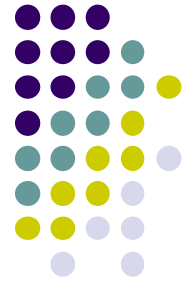




# Remarks

1. Theoretically, the proposal distribution can be arbitrary; however, it is wise to choose a standard distribution which (i) approximates to the target distribution; (ii) can be sampled from easily.
2. When the proposal distribution is symmetric with respect to  $X_{\text{can}}$  and  $X^{(l)}$ , the acceptance probability may reduce to

$$p_{\text{accept}} = \min \left\{ 1, \frac{f(X_{\text{can}})}{f(X^{(l)})} \right\}.$$



# Remarks

3. The proposal distribution may include a tuning parameter that controls the acceptance rate.
4. The optimal acceptance rate is chosen on the problem-by-problem basis (Gelman et al. 1995).

- Reference

- Gelman, A., Roberts, G. O. and Gilks, W. R. Efficient Metropolis jumping rules. In J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith (Eds), *Bayesian Statistics 5*, pp. 599-607. Oxford: Oxford University Press, 1995.

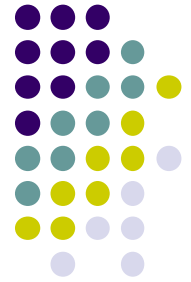


## Example: Nonlinear SEM

- Target distribution

$$\exp \left\{ -\frac{1}{2} \boldsymbol{\xi}_i^T \boldsymbol{\Phi}^{-1} \boldsymbol{\xi}_i - \frac{1}{2} (\mathbf{y}_i - \mathbf{A} \mathbf{c}_i - \boldsymbol{\Lambda} \boldsymbol{\omega}_i)^T \boldsymbol{\Psi}_{\epsilon}^{-1} (\mathbf{y}_i - \mathbf{A} \mathbf{c}_i - \boldsymbol{\Lambda} \boldsymbol{\omega}_i) \right. \\ \left. - \frac{1}{2} (\boldsymbol{\eta}_i - \mathbf{B} \mathbf{d}_i - \boldsymbol{\Pi} \boldsymbol{\eta}_i - \boldsymbol{\Gamma} \mathbf{F}(\boldsymbol{\xi}_i))^T \boldsymbol{\Psi}_{\delta}^{-1} (\boldsymbol{\eta}_i - \mathbf{B} \mathbf{d}_i - \boldsymbol{\Pi} \boldsymbol{\eta}_i - \boldsymbol{\Gamma} \mathbf{F}(\boldsymbol{\xi}_i)) \right\}.$$

Proposal distribution:  $N(\boldsymbol{\omega}_i^{(l)}, \sigma_\omega^2 \boldsymbol{\Sigma}_\omega)$



1.  $\boldsymbol{\Sigma}_\omega^{-1} = \boldsymbol{\Sigma}_\delta^{-1} + \boldsymbol{\Lambda}^T \boldsymbol{\Psi}_\epsilon^{-1} \boldsymbol{\Lambda}.$

2.  $\boldsymbol{\Sigma}_\delta^{-1} = \begin{bmatrix} \boldsymbol{\Pi}_0^T \boldsymbol{\Psi}_\delta^{-1} \boldsymbol{\Pi}_0 & -\boldsymbol{\Pi}_0^T \boldsymbol{\Psi}_\delta^{-1} \boldsymbol{\Gamma} \boldsymbol{\Delta} \\ \text{symmetric} & \boldsymbol{\Phi}^{-1} + \boldsymbol{\Delta}^T \boldsymbol{\Gamma}^T \boldsymbol{\Psi}_\delta^{-1} \boldsymbol{\Gamma} \boldsymbol{\Delta} \end{bmatrix}.$

3.  $\boldsymbol{\Pi}_0 = \boldsymbol{I} - \boldsymbol{\Pi}.$

4.  $\boldsymbol{\Delta}^T = \left. \frac{\partial F(\xi_i)}{\partial \xi_i} \right|_{\xi_i=0}.$



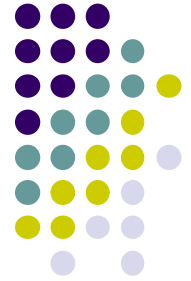
## Remarks

1.  $\Sigma_{\omega}^{-1}$  is the Hessian matrix of  $-\log f(\omega_i | \mathbf{y}_i, \theta)$  with  $\omega_i = \mathbf{0}$ .
2.  $\sigma_{\omega}^2$  is the tuning parameter which is chosen such that the average acceptance rate is approximately 0.25 or more.



# More References

- Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. *Bayesian Data Analysis* (2<sup>nd</sup> Edition). Chapman and Hall, 2004.
- Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (Eds) *Markov Chain Monte Carlo in Practice*. Chapman and Hall, 1996.
- Robert, C. P. and Casella, G. *Monte Carlo Statistical Methods* (2<sup>nd</sup> Edition). Springer, 2010.



**Thank you!**