CHAPTER 16
MARKOV CHAIN MONTE CARLO

• Organization of chapter in ISSO
  – Background on MCMC
  – Metropolis-Hastings algorithm
  – Numerical example of Metropolis-Hastings
  – Gibbs sampling
  – Numerical example of Gibbs sampling
  – Optional in these slides: Non-Gaussian state estimation (not in ISSO)
Background

- Process generating random vector $X$
- Want to compute $E([f(X)]$ for function $f(\cdot)$
- Standard method for approximating $E([f(X)]$ is to generate many *independent* sample values of $X$ and compute sample mean of $f(X)$ values
- Only useful in “trivial” cases where $X$ can be generated directly
- Many practical problems have non-trivial distribution for $X$
  - E.g., state in nonlinear/non-Gaussian state-space model
Markov Chains

• Not necessary to generate independent $X$ to estimate $E[f(X)]$
• Consider \textit{dependent} sequence $X_0, X_1, X_2, \ldots$
• Generate $X_{k+1}$ according to “easy” conditional distribution for $\{X_{k+1}|X_k\}$
  – $\{X_k\}$ process is a Markov chain
  – $X_k$ dependence on fixed number of early states disappears as $k$ gets large
• Above implies distribution of $X_k$ approaches a stationary form as $k$ gets large
  – Stationary form corresponds to \textit{target distribution} (density) $p(\cdot)$ if conditional distribution chosen properly
Ergodic Averaging

• Let $M$ denote the “burn-in” period for the Markov chain

• The ergodic average of $n - M$ values of $f(X)$ with $X_k$ generated via a Markov chain is

\[
\frac{1}{n-M} \sum_{k=M+1}^{n} f(X_k)
\]

• Summands above are dependent via the Markov property for the $\{X_k\}$

• Above sum approaches $E([f(X)]$ as $n$ gets large by ergodic theorem
Metropolis-Hastings (M-H) Algorithm

• M-H algorithm is one of two most popular forms for MCMC (other is Gibbs sampling)
• M-H relies on proposal distribution and Metropolis criterion
• Let proposal distribution be \( q(\cdot|\cdot) \); used to generate candidate points \( W \sim q(\cdot|X=x) \)
• Candidate point \( W = w \) is accepted with probability given by Metropolis criterion:

\[
\rho(x, w) \equiv \min \left\{ \frac{\rho(w) q(x|w)}{\rho(x) q(w|x)}, 1 \right\}
\]

• In practice, in going from \( X_k \) to \( X_{k+1} \), \( x \) above is \( X_k \) and \( W \) becomes \( X_{k+1} \) if \( W \) is accepted
M-H Algorithm for Estimating $E([f(X)])$

**Step 0 (initialization)** Choose length of “burn-in” period $M$ and initial state $X_0$. Set $k = 0$.

**Step 1 (candidate point)** Generate a candidate point $W$ according to proposal distribution $q(\cdot | X_k)$.

**Step 2 (accept/reject)** Generate point $U$ from $U(0, 1)$ distribution. Set $X_{k+1} = W$ if $U \leq \rho(X_k, W)$ (Metropolis criterion). Otherwise set $X_{k+1} = X_k$.

**Step 3 (iterate)** Repeat Steps 1 and 2 until $X_M$ is available. Terminate “burn-in” process and proceed to step 4 with $X_k = X_M$.

**Steps 4–6 (ergodic average)** Repeat process and compute average of $f(X_{M+1}), \ldots, f(X_n)$. This ergodic average is estimate of $E([f(X)])$. 

16-6
Example: Estimating $E([f(X)])$ from a Bivariate Normal Distribution (Example 16.1 from ISSO)

- Suppose $X \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix} \right)$

- Use M-H to estimate sum of the two mean components (true value = 0): $f(X) = [1, 1]X$

- Standard (unit length) uniform proposal distribution and burn-in period of $M = 500$

- Following plot shows three independent runs
  - Acceptance rate (Metropolis criterion) about 70%
  - Better performance possible with lower acceptance rate (requires “tuning”—not always feasible in practice)
Example (cont’d): M-H Algorithm with Uniform Proposal Distribution; Mean Zero Target