MCMC Methods

Feng Li

feng.li@cufe.edu.cn

School of Statistics and Mathematics
Central University of Finance and Economics

Outline

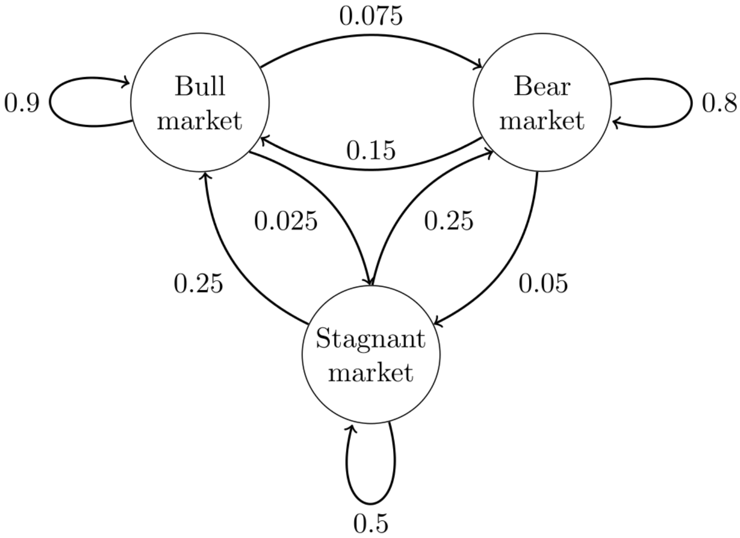
Markov chain

- A **Markov chain** is a sequence of random variables $X_1, X_2, X_3, \ldots$ with the **Markov property**, namely that, given the present state, the future and past states are independent. Formally,

$$
\begin{aligned}
&\Pr(X_{n+1} = x | X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) \\
&= \Pr(X_{n+1} = x | X_n = x_n).
\end{aligned}
$$

- **Example**: **hypothetical stock market**

Hypothetical stock market example



Hypothetical stock market example

- The **transition matrix** for this example is

$$P = \begin{bmatrix} 0.9 & 0.075 & 0.025 \\ 0.15 & 0.8 & 0.05 \\ 0.25 & 0.25 & 0.5 \end{bmatrix}.$$

- The distribution over states can be written as a stochastic row vector x with the relation $x(n+1) = x(n)P$.

$$\begin{aligned} x^{(n+3)} &= x^{(n+2)}P = \left(x^{(n+1)}P\right)P = x^{(n+1)}P^2 = \left(x^{(n)}P^2\right)P \\ &= x^{(n)}P^3 = \begin{bmatrix} 0.3575 & 0.56825 & 0.07425 \end{bmatrix}. \end{aligned}$$

- Furthermore each column of

$$\lim_{N\to\infty} P^N = \begin{bmatrix} 0.625 & 0.3125 & 0.0625 \\ 0.625 & 0.3125 & 0.0625 \\ 0.625 & 0.3125 & 0.0625 \end{bmatrix}$$

is the **stationary distribution**.

Markov chain Monte Carlo (MCMC)

- **Markov chain Monte Carlo**: to simulate from a distribution π (for instance, the posterior distribution), it is actually sufficient to produce a Markov chain X_t where $t \in N$ whose **stationary distribution** is π
- If an algorithm that generates such a chain can be constructed, the ergodic theorem guarantees that, in almost all settings, the average

$$\frac{1}{T}\sum_{t=1}^{T} g(x_t)$$

  converges to $E(g(x))$ no matter what the starting value is.
- Gibbs sampler is an MCMC algorithm.
- Metropolis-Hastings is an MCMC algorithm.

Diagnosing Convergence

- We need a **stopping rule** to guarantee that the number of iterations is sufficient.
- Criteria
  - Convergence to the Stationary Distribution
  - Convergence of Averages
  - Convergence to iid Sampling

Convergence in multiple chains

- Many multiple-chain convergence diagnostics are quite elaborate.
- The performances of these parallel methods require a degree of a priori knowledge on the distribution in order to construct an initial distribution.
  - An initial distribution which is too concentrated around a local mode does not contribute significantly more than a single chain to the exploration
  - Moreover, slow algorithms, Gibbs sampling used in highly nonlinear setups, usually favor single chains.
- It is somewhat of an illusion to think we can control the flow of a Markov chain and assess its convergence behavior from a few realizations of this chain.

Monitoring Convergence of Distribution

- A natural empirical approach to convergence control is to draw pictures of the output of simulated chains, in order to detect deviant or non-stationary behaviors. However, this plot is only useful for strong non-stationarities of the chain.
- Tests for non-stationary checking.
  - Autocorrelation functions.
- Another approach to convergence monitoring is to assess how much of the support of the target distribution has been explored by the chain via an evaluation of

$$\int_A f(x)dx \approx \sum_{t=1}^{T-1} (\theta(t+1) - \theta_t) f(\theta_t)$$

when $f(x)$ is a one-dimensional density, the above converges to 1.

Monitoring Convergence of Average

- Graphical outputs can detect obvious problems of convergence of the empirical average.
- One may use cumulative sums (CUSUM), graphing the partial differences

$$D_T^* \sum_{t=1}^{i} \left(h(\theta_t) - \frac{1}{T}\sum_{t=1}^{T} h(\theta_t)\right)$$

Effective Sample Size

- The standard approach to restore to the **effective sample size** which gives the size of an iid sample with the same variance as the current sample and thus indicates the loss in efficiency due to the use of a Markov chain. This value is computed as

$$\frac{T}{1 + 2\sum_{t=1}^{\infty} \mathrm{Corr}(h(\theta_0), h(\theta_t))}$$

where the denominator is the measurement of efficiency (**inefficiency factor**)

Further Suggested Read

Monte Carlo Statistical Methods Book by Christian P Robert and George Casella. (2004 edition)