# PRML chapter11

1. Markov Chain Monte Carlo
   - allows sampling from a large class of distributions, and which scales well with the dimensionality of the sample space
   - Metropolis algorithm
     - sample is accepted with probability

$$A\left(\mathbf{z}^{\star}, \mathbf{z}^{(\tau)}\right) = \min\left(1, \frac{\widetilde{p}\left(\mathbf{z}^{\star}\right)}{\widetilde{p}\left(\mathbf{z}^{(\tau)}\right)}\right)$$

     - choose a random number u with uniform distribution over the unit interval (0,1)
     - accepting the sample if $A\left(\mathbf{z}^{\star}, \mathbf{z}^{(\tau)}\right) > u$
     - a central goal in designing Markov chain Monte Carlo methods is to avoid random walk behaviour

   (a) Markov chains
   - Markov chain is the situation that we can approximate future from the given states
   - A first-order Markov chain is defined as

$$p\left(\mathbf{z}^{(m+1)}|\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(m)}\right) = p\left(\mathbf{z}^{(m+1)}|\mathbf{z}^{(m)}\right)$$

   - marginal probability is written as

$$p\left(\mathbf{z}^{(m+1)}\right) = \sum_{\mathbf{z}(m)} p\left(\mathbf{z}^{(m+1)}|\mathbf{z}^{(m)}\right) p\left(\mathbf{z}^{(m)}\right)$$

   - A distribution is invariant when it satisfies

$$p^{\star}(\mathbf{z}) = \sum_{\mathbf{z}'} T\left(\mathbf{z}', \mathbf{z}\right) p^{\star}\left(\mathbf{z}'\right)$$

   (b) The Metropolis-Hastings algorithm
   - We first propose q(Z), then samples from it. We accept the sample when it sattisfies detailed balance
   - generalize metropolis algorithm by making

$$A_k\left(\mathbf{z}^{\star}, \mathbf{z}^{(\tau)}\right) = \min\left(1, \frac{\widetilde{p}\left(\mathbf{z}^{\star}\right) q_k\left(\mathbf{z}^{(\tau)}|\mathbf{z}^{\star}\right)}{\widetilde{p}\left(\mathbf{z}^{(\tau)}\right) q_k\left(\mathbf{z}^{\star}|\mathbf{z}^{(\tau)}\right)}\right)$$

   - By using this we could proof the Metropolis algorithm samples from the required distribution

2. Gibbs Sampling
   - a special case of the Metropolis- Hastings algorithm.
   - First we initialize $\{z_i : i = 1, \dots, M\}$
   - for each $\tau$ :and for each j: sample

$$z_j^{(\tau+1)} \sim p\left(z_j | z_1^{(\tau+1)}, \dots, z_{j-1}^{(\tau+1)}, z_{j+1}^{(\tau)}, \dots, z_M^{(\tau)}\right)$$

   - In order to gain the proper sampling, we need $p(z)$ is invariant and Ergodicity
   - we use "over-relaxation" to prevent it behaving like random walk.
3. Slice Sampling
   - Metropolis algorithm is sensitive to step size
   - The technique of slice sampling provides an adaptive step size that is automatically adjusted to match the characteristics of the distribution
4. The Hybrid Monte Carlo Algorithm
   (a) Dynamical systems
      - momentum variable

$$r_i = \frac{\mathrm{d}z_i}{\mathrm{d}\tau}$$

      - under $E(z)$ is potential energy

$$p(\mathbf{z}) = \frac{1}{Z_p} \exp(-E(\mathbf{z}))$$

      - Physical energy is

$$K(\mathbf{r}) = \frac{1}{2}\|\mathbf{r}\|^2 = \frac{1}{2}\sum_i r_i^2$$

      - The total enegy of the system is

$$H(\mathbf{z}, \mathbf{r}) = E(\mathbf{z}) + K(\mathbf{r})$$

      - During the evolution of this dynamical system, Hamilton function is constant
      - Consider the joint distribution over phase space whose total energy is the Hamiltonian

$$p(\mathbf{z}, \mathbf{r}) = \frac{1}{Z_H} \exp(-H(\mathbf{z}, \mathbf{r}))$$

      - leapfrog discretization
         – repeat following loop

$$\widehat{r}_i(\tau + \epsilon/2) = \widehat{r}_i(\tau) - \frac{\epsilon}{2}\frac{\partial E}{\partial z_i}(\widehat{\mathbf{z}}(\tau))$$

$$\widehat{z}_i(\tau + \epsilon) = \widehat{z}_i(\tau) + \epsilon \widehat{r}_i(\tau + \epsilon/2)$$

$$\widehat{r}_i(\tau + \epsilon) = \widehat{r}_i(\tau + \epsilon/2) - \frac{\epsilon}{2}\frac{\partial E}{\partial z_i}(\widehat{\mathbf{z}}(\tau + \epsilon))$$

   (b) Hybrid Monte Carlo
      - combine Hamilton dynamics and Metropolis algorithm
   (c) Estimating the Partition Function

- in order to compare Bayes models, need to know the odds of $Z_E$
- $$\frac{Z_E}{Z_G} = \frac{\sum_{\mathbf{z}} \exp(-E(\mathbf{z}))}{\sum_{\mathbf{z}} \exp(-G(\mathbf{z}))} = \frac{\sum_{\boldsymbol{x}} \exp(-E(\mathbf{z}) + G(\mathbf{z})) \exp(-G(\mathbf{z}))}{\sum_{\mathbf{z}} \exp(-G(\mathbf{z}))} = E_{G(\mathbf{z})}[\exp(-E + G)]$$
  $$\simeq \frac{1}{L} \sum_{l} \exp\left(-E\left(\mathbf{z}^{(l)}\right) + G\left(\mathbf{z}^{(l)}\right)\right)$$