Stellenbosch University Iyunivesithi Universiteit
Department of Mathematical Sciences
Division of Applied Mathematics

AIMS | Physics for Machine Learning

Problems for Chapter 2: Markov chain Monte Carlo

Reading

- ML822 Monte Carlo notes: whole of Chap. 2.
- AM783 Markov Processes notes: whole of Chap. 2.

Numerical

Q1. MCMC estimation of π. Estimate π as done in CW1, replacing the independent points by a random walk in the square. You can use symmetric displacements for the random walk.

(a) Show numerically that the points generated are uniform in the square.

(b) Show numerically that your estimator of π converges to the correct value. Report your acceptance ratio. [Hint: Use a "good" distribution for the displacements.]

(c) Can we construct error bars for this estimator the way we did in CW1? Explain your answer.

Q2. Gaussian mixture. Use the Metropolis algorithm to sample from the following Gaussian mixture:

$$p(x) = a_1 p_1(x) + a_2 p_2(x), \tag{1}$$

where $p_1 = \mathcal{N}(-20, 100)$, $p_2 = \mathcal{N}(20, 100)$ (so standard deviation = 10), $a_1 = 0.3$, and $a_2 = 0.7$. Consider Gaussian displacements with standard deviation 1, 500 and 8. Discuss your results for each.

Q3. Bayesian linear fit. Our goal in this problem is to use MCMC to sample the distribution $p(\theta|D)$ of the parameters θ of some model given data D from that model. The model that we consider is a linear model perturbed by Gaussian noise:

$$y = ax + b + \sigma z$$

where $z \sim \mathcal{N}(0, 1)$. We thus have 3 parameters $\theta = (a, b, \sigma)$.

(a) Generate 50 data points from this model using $a = 1$, $b = 0$, and $\sigma = 1$ (the real, hidden parameter values). This is your (fixed) data D. Plot the data along with a least-square linear fit.

(b) What's the expression of the likelihood $p(D|\theta)$ for this model? From this distribution, obtain an explicit expression for the posterior $p(\theta|D)$ using Bayes's formula and the following prior:

$$p(\theta) = p(a)p(b)p(\sigma),$$

where $p(a)$ and $p(b)$ and $p(\sigma)$ are all assumed to be Gaussian distributions with hyper-parameters that you need to choose or determine from the data. When writing $p(\theta|D)$, keep the evidence $p(D)$ indeterminate – we don't have an expression for this probability (yet).

(c) Construct an MCMC algorithm for sampling the posterior $p(\theta|D)$ based on the unnormalised expression of this distribution found in (b). Use the following steps in your code:

(i) Propose Gaussian moves for a, b, and σ with mean 0 and appropriate small variance.

(ii) Accept moves with probability

$$\rho = \min\left\{1, \frac{p(\theta'|D)}{p(\theta|D)}\right\} = \min\left\{1, \frac{p(D|\theta')p(\theta')}{p(D|\theta)p(\theta)}\right\}.$$

(iii) Use the log of probabilities everywhere to avoid getting too large or too small values.

(iv) Track the acceptance ratio in your simulation and change the moves' variance to reach a ratio close to 0.5.

(d) Plot histograms of sampled values for a, b and σ, representing the marginals of $p(\theta|D)$. From the results, determine the most probable values of the parameters for the data that you have.