Lecture Summary 5

Most of the sampling algorithms need to return to start after a single sample has been found. It would be a nice idea if it were possible to use this sample to help find a next one. This is the idea behind adaptive rejection sampling where the proposal density can be improved. However, this is restricted to log concave density functions. There are other ways to adapt a proposal density but we will not look at them now.

A sensible idea is to look for samples close to an accepted sample. So suppose we have $X_1 \sim f$ and are looking to get another $X_2$. Instead of sampling $f$ suppose we sample a conditional density $p(x|X_1)$. We need to find a suitable $p(x|X_1)$, which we will use for every sample, so to get $X_{n+1}$ we sample from $p(x|X_n)$, such that

$$\widehat{I}_N = N^{-1} \sum_{i=1}^{N} l(X_i) \to I = \int l(x)\, f(x)\, dx.$$

There are such $p$, which satisfy

$$f(X_{n+1}) = \int p(X_{n+1}|X_n)\, f(X_n)\, dX_n$$

for all $n$. It should be much easier to sample $p(x|X_n)$ than $f(x)$ provided it is chosen carefully.

There is a drawback, which is that the variance of $\widehat{I}_N$ is more than it is from an estimator based on an i.i.d. sample from $f$. In fact

$$\sqrt{N}(\widehat{I}_N - I) \to N(0, \sigma^2)$$

where

$$\sigma^2 = \operatorname{Var} g(X_1) + 2 \sum_{i=1}^{\infty} \operatorname{Cov}(g(X_1), g(X_i))$$

where $X_1 \sim f$ and the covariance is with respect to the Markov sample. So the penalty to pay is the extra covariance terms.

So we need to find a $p(x|X_n)$ which we can sample easily and produces not too high a covariance.

We will look at a Markov model which would not be needed in practice but does illustrate the ideas. So suppose

$$p(x|X_n) = N\Big(\rho X_n, \sqrt{1-\rho^2}\Big),$$

for some $-1 < \rho < 1$. This gives rise to a Markov sequence $(X_n)$ which we will study further in class.