Statistical Models and Computing Methods, Problem Set 2

October 27, 2021 Due 11/11/2021

Problem 1.

Consider estimating the following integral

$$I = \int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2}\right) dx$$

using importance sampling.

(1) Implement importance sampling with the Laplace distribution $q(x) = \frac{1}{2}\exp(-|x|)$ as the importance distribution. Report the mean and variance of your estimate as a function of the number of samples.

(2) Now use the family of exponential distributions $q_\lambda(x) = \frac{\lambda}{2}\exp(-\lambda|x|)$ as candidate distributions for adaptive important sampling. Find the best λ using both the variance minimization and cross entropy minimization criteria. Report the variances of your estimates as a function of the number of iterations for both methods. You may estimate the variances using 1000 samples. Do you see any difference between the training processes of the two methods? Explain it.

Problem 2.

If $K(x, x')$ is a Markov transition kernel with stationary distribution g, show that the Metropolis-Hastings algorithm where, at iteration t, the proposed value $y_t \sim K(x^{(t)}, y)$ is accepted with probability

$$\min\left(1, \frac{f(y_t)}{f(x^{(t)})}\frac{g(x^{(t)})}{g(y_t)}\right)$$

provide a valid MCMC algorithm for the stationary distribution f.

Problem 3.

Consider the probit regression model, with $x_i \in \mathbb{R}$ and $Y_i \in \{0, 1\}$:

$$\begin{aligned} p(Y_i = 1|Z_i) &= p(Z_i \geq 0) \\ Z_i &\sim \mathcal{N}(x_i\beta, \sigma^2) \\ \beta &\sim \mathcal{N}(0, 10^2) \\ \sigma^2 &\sim \text{Inv-}\chi^2(3, 1), \end{aligned}$$

where $(x_i, y_i),\ i = 1, \ldots, n$ are the observe data. Download the data from the course website.

(1) Implement a Gibbs sampler for approximating the posterior of (β, σ^2). Initialize the sampler at $(25, 25)$ and run it for $20,000$ iterations (with no burn-in). Plot the samples on top of the contours of the true posterior.
(2) Now consider an alternative sampler that inserts a Metropolis-Hasting step after each Gibbs cycle which scales the current state of the Markov chain, $(\beta^{(t)}, (\sigma^2)^{(t)})$, by a factor s which is drawn from an exponential distribution, Exp(1). The rescaled state is accepted or rejected according to the usual Metropolis-Hastings procedure. Implement this sampler and conduct a similar simulation as in part (1), again plotting the samples on top of the contours of the true posterior. Which sampler mixes more quickly?

Problem 4.
Consider a logistic regression model with normal priors

$$y_i \sim \text{Bernoulli}(p_i),\ p_i = \frac{1}{1 + \exp(-x_i^T \beta)},\quad i = 1, \ldots, n. \quad \beta \sim \mathcal{N}(0, \sigma_\beta^2)$$

where $\sigma_\beta = 1$. Download the data from the course website.

(1) Implement a Hamiltonian Monte Carlo sampler to collect 500 samples (with 500 discarded as burn-in), show the scatter plot. Test the following two strategies for the number of leapfrog steps L: (1) use a fixed L; (2) use a random one, say Uniform$(1, L_{\max})$. Do you find any difference? Explain it.
(2) Run HMC for 100000 iterations and discard the first 50000 samples as burn-in to form the ground truth. Implement stochastic gradient MCMC algorithms including SGLD, SGHMC and SGNHT. Show the convergence rate of different SGMCMC algorithms in terms of KL divergence to the ground truth as a function of iterations. You may want to use the ITE package `https://bitbucket.org/szzoli/ite-in-python/src/` to compute the KL divergence between two samples.