High Performance Computing for Science and Engineering II

Spring semester 2019

P. Koumoutsakos
ETH Zentrum, CLT E 13
CH-8092 Zürich

HW 3 (Part 1 of 2): Sampling Methods

Issued: March 18, 2019
Due Date: April 1, 2019 10:00am

Task 1: Inversion Method for Gaussian Sampling (20 Points)

Consider the random variable

$$X = \sqrt{2}\,\mathrm{erf}^{-1}\left(2U - 1\right), \text{ where } U \sim \mathcal{U}(0,1).$$

The error function is defined as

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2)\,\mathrm{d}t.$$

a) (15 Points) Use the inversion method to show that $X \sim \mathcal{N}(0,1)$.

b) (5 Points) Explain why the above formula is not used to generate normally distributed random numbers in practice (2 lines).

Task 2: Importance Sampling (30 Points)

Taking advantage of the importance sampling method, we can rewrite the integral

$$\mathbb{E}_f[h(X)] = \int h(x)\, f(x)\, dx, \tag{1}$$

into

$$\mathbb{E}_g\Big[h(X)\frac{f(X)}{g(X)}\Big] = \int h(x)\, \frac{f(x)}{g(x)}\, g(x)\, dx\,, \tag{2}$$

for any density g that satisfies $\mathrm{supp}(f) \subseteq \mathrm{supp}(g)$. Then the estimator for Eq.1:

$$\hat{I} = \frac{1}{N}\sum_{i=1}^{N} h(Y_i)\,, \tag{3}$$

,where Y_i are i.i.d. samples from f, turns into the estimator for Eq.2

$$\hat{I} = \frac{1}{N}\sum_{i=1}^{N} h(X_i)\, \frac{f(X_i)}{g(X_i)}\,, \tag{4}$$

where X_i are i.i.d. samples from g.

a) (15 Points) Show that the optimal choice of g, in the sense that it minimizes the variance of estimator (Eq.4), is given by

$$g^\star(x) = \frac{|h(x)|f(x)}{\int |h(x)|f(x)\, dx}\,. \tag{5}$$

Can you argue why this result is not in practice employed?
(hint: use Jensen's inequality: In the context of probability theory, if X is a random variable and ϕ *is a convex function, then:* $\phi(\mathbb{E}[\mathrm{X}]) \le \mathbb{E}[\phi(\mathrm{X})]$*).*

b) (15 Points) Write a program to estimate the probability $P(X > 4.5)$ for $X \sim \mathcal{N}(0, 1)$ using:

1. the estimator (3), with $h \equiv 1$, if $(X > 4.5)$ and $h \equiv 0$ otherwise.
2. the estimator (4), where $h \equiv 1$, if $(X > 4.5)$ and $h \equiv 0$ otherwise, f the density of the normal distribution and g the density of an exponential distribution, truncated at 4.5 with scale 1.

Compare your results with the exact value that can be obtained using the cumulative distribution function of the normal distribution. What do you observe for $N = 10^4$? Can you argue on the performance of the two estimators?

Task 3: MCMC: Hastings and application to coin-toss problem (40 Points)

In his 1970's paper Hastings proposed a general form of the acceptance probability $\alpha(x|y)$ for MCMC algorithms,

$$\alpha(x|y) = \frac{s(x|y)}{1 + \frac{q(x|y)p(y)}{q(y|x)p(x)}}, \tag{6}$$

where q represents the proposal distribution (in general non-symmetric), p the stationary distribution, and s any symmetric function ($s(x|y) = s(y|x)$) which can guarantee that $\alpha(x|y) \leq 1$ for all x, y.

a) (10 Points) Show that the transition probability $t(x|y) = \alpha(x|y)q(x|y)$ with α chosen as above satisfies the *detailed balance*. Note that you do not have to prove that the expression for the transition probability holds.

b) (10 Points) Show that the Metropolis choice: $\alpha_M(x|y) = \min\{1, \frac{q(y|x)p(x)}{q(x|y)p(y)}\}$ is a special case of the above formula. What is $s(x|y)$ in this case?

c) (5 Points) It is often a good practice to convert your calculations for the MCMC algorithm in the logarithmic scale. Write down how the formulation of one step of the Metropolis-Hastings algorithm changes, if we convert all probabilities to the log-scale. Can you think of a reason why this formulation can be advantageous for numerical applications?

d) (10 Points) We provide you with a **skeleton code** that implements MCMC based on the Metropolis-Hastings rejection criterion in order to sample from the posterior distribution $p(H|\boldsymbol{d})$ from the coin flip example in Section 2.2 of the Lecture Notes. Recall that from the Bayes theorem: $p(H|\boldsymbol{d}) = p(\boldsymbol{d}|H)p(H)$. As a proposal distribution from moving from state x to state y, $q(y|x)$, we use the **symmetric** Gaussian distribution with mean $\mu = 0$ and standard deviation $\sigma = 0.1$.
You are asked to implement the version using logarithmic scale of the probabilities involved in the provided code. Fill in the "TODO"s in the provided code.
Run your code and plot a histogram of both the samples drawn from the original MCMC and the modified MCMC in the logarithmic scale. Compare the two results.

e) (5 Points) Change the number of tosses to 3000 and the number of heads to 1500. Try to plot a histogram of the samples drawn with the original MCMC and the modified MCMC in logarithmic scale. What do you observe?

Guidelines for reports submissions:

- Submit a report in .pdf format and the two solution codes (recommended: in python) by Monday 01.04.
- The final version of your report should include the solutions to the second part of Homework 3, which will be handed out on Monday 25.03.2019.