Boğaziçi University, Dept. of Computer Engineering

CMPE 58N, MONTE CARLO METHODS

Spring 2009, Midterm

Name: ______________________________

Student ID: __________________________

Signature: ___________________________

- Please print your name and student ID number and write your signature to indicate that you accept the University honour code.
- During this examination, you may use any notes, books or laptops. You can even lookup resources on the internet; however communication with fellow students is not allowed.
- Read each question carefully and show all your work. Underline your final answer to each question.
- There are **7** questions. Point values are given in parentheses.
- You have **180 minutes** to do all the problems.

Q	1	2	3	4	5	6	7	Total
Score								
Max	8	12	12	10	10	12	16	80

1. (**What is ...**) Use the space below each question. Give concise answers, long answers ($>$ 2 sentences) don't get any points.

   (a) *(1 pts)* What is a Gibbs sampler?

   A particular MH algorithm where we sample from full conditionals

   (b) *(1 pts)* What is ergodicity?

   For Markov chains, aperiodicity **and** irreducibility. For sufficiently large times, the chain can visit every state from every state at any time.

   (c) *(1 pts)* In Monte Carlo estimation, how does the error "behave" with increasing sample size?

   $N^{-1/2}$

   (d) *(1 pts)* What does the central limit theorem say about Monte Carlo integration?

   The variance diminishes with incresing sample size

   (e) *(1 pts)* Can one use the inversion method to generate Gaussian random variables ? Why (not) ?

   The CDF of a Gaussian is not known in closed form. Hence, the inversion method is not practical.

   (f) *(1 pts)* When are the estimates obtained by importance sampling biased ?

   When the normalisation constant is unknown

(g) *(2 pts)* Show that when the kernel T_1 and T_2 have the same unique invariant distribution, their mixture has the same invariant distribution.

Call the invariant distribution π. We have

$$T_1\pi = \pi$$
$$T_2\pi = \pi$$

Define a mixture Kernel

$$T = \nu T_1 + (1-\nu)T_2$$

$$\begin{aligned} T\pi &= \nu T_1\pi + (1-\nu)T_2\pi \\ &= \nu\pi + (1-\nu)\pi = \pi \end{aligned}$$

2. (**Quiz question**) Given

$$g_i \sim \mathcal{G}(g_i; a_i, \lambda)$$

for $i = 1, 2, 3$, find the distribution of

$$z = \frac{g_1 + g_2}{g_1 + g_2 + g_3}$$

(12 points) Show that $g = g_1 + g_2 \sim \mathcal{G}(g; a_1 + a_2, \lambda)$. The rest is similar to the quiz solution

(blank page for further space)

3. **(Random bit source)** Suppose we have a source of random bits A where each bit a_i is independent Bernoulli with **unknown** probability q. Using this source, devise an algorithm to generate uniform Bernoulli bits b_j, each with probability p. Find the efficiency of your approach, i.e., the expected number of true random bits b per random bits a, as a function of p and q. *(12 points)*

   Obvious solution:

   (a) Use the a_i to generate c_k where $c_k \sim \mathcal{B}(c_k; 1/2)$
   (b) Use the c_k to generate b_j

$$c_k = \begin{cases} 0 & \text{if } a_{2i-1} = 0 \text{ and } a_{2i} = 1 \\ 1 & \text{if } a_{2i-1} = 1 \text{ and } a_{2i} = 0 \\ \text{reject} & \end{cases}$$

   Note that we consider non-overlapping pairs. Otherwise, the trials may not be independent (need to check this).

   The efficiency of this approach is as follows:

   The probability of a rejection is:

$$\Pr\{\text{reject}\} = (1-q)^2 + q^2 = 1 + 2q^2 - 2q$$

   and

$$\Pr\{\text{accept}\} = 2q - 2q^2$$

   Hence, the expected number of bits c per 2 draws from A is

$$\langle N \rangle = \Pr\{\text{accept}\}1 + \Pr\{\text{reject}\}0 = \Pr\{\text{accept}\}$$

   Consequently, we expect to get $q(1-q)$ bits per draw from A.

   Clearly, we should be able to think of better ways. For example, the maximum efficiency is achieved when $q = 0.5$ where we loose 3 bits out of 4.

   Now, since we have unbiased bits c, we can use the Knuth and Yao algorithm.

   The efficiency of KY is easy. Given any binary bitstream that defines $p = 0.p_1p_2\ldots$, the probability to terminate the loop is $z = 0.5$. Hence the expected number of times we execute the loop is

$$0.5 \times 1 + 0.5^2 \times 2 + 0.5^3 \times 3 \cdots = 2$$

   This can be seen as follows: For $0 \le z < 1$

$$\sum_{k=0}^{\infty} z^k = \frac{1}{1-z}$$

$$\frac{d}{dz}\sum_{k=0} z^k = \frac{d}{dz}\left(\frac{1}{1-z}\right)$$

$$\sum_{k=1}^{\infty} kz^{k-1} = 1/(1-z)^2$$

$$\sum_{k=1}^{\infty} kz^k = z/(1-z)^2$$

Evaluate for $z = 0.5$.

As steps 1 and 2 are independent. The expected number of bits is about

$$q(1-q)/2$$

It would be nice to implement this algorithm and verify this result via simulation.

4. **(Rejection Sampling)** Consider a symmetric triangle distribution p on $[-1, 1]$. Using a uniform distribution as the proposal q, compute the efficiency of rejection sampling, i.e., how many samples from p can be generated per sample from q?

(10 points)

5. **(Importance sampling)** Consider the target distribution

$$p(x) = \mathcal{N}(x; m, S)$$

when using the proposal

$$q(x) = \mathcal{N}(x; \mu, \Sigma)$$

(a) Derive the weight function for the target distribution

(b) Derive the analytic expression for the variance of importance weights

(10 points)

(a) Weight function:

$$\begin{aligned}
p(x) &= \mathcal{N}(x; m, S) = \exp(-\frac{1}{2S}(x-m)^2 - \frac{1}{2}\log(2\pi S)) \\
q(x) &= \mathcal{N}(x; \mu, \Sigma) = \exp(-\frac{1}{2\Sigma}(x-\mu)^2 - \frac{1}{2}\log(2\pi \Sigma)) \\
W(x) &= p(x)/q(x) = \exp(-\frac{1}{2S}(x-m)^2 + \frac{1}{2\Sigma}(x-\mu)^2 - \frac{1}{2}\log(S/\Sigma)) \\
&= \exp(-\frac{1}{2}\left(\frac{1}{S} - \frac{1}{\Sigma}\right)x^2 - \frac{1}{2}\left(\frac{1}{S} - \frac{1}{\Sigma}\right)m^2 + \left(\frac{1}{S}m - \frac{1}{\Sigma}\mu\right)x - \frac{1}{2}\log(S/\Sigma))
\end{aligned}$$

This is proportional to a Gaussian with

$$\begin{aligned}
W(x) &\propto \mathcal{N}(x; m_W, S_w) \\
S_w &= \left(\frac{1}{S} - \frac{1}{\Sigma}\right)^{-1} \\
m_W &= S_w\left(\frac{1}{S}m - \frac{1}{\Sigma}\mu\right)
\end{aligned}$$

Note that if $\Sigma < S$, the variance becomes negative; hence the Gaussian becomes unnormalisable.

(b) Variance

$$\begin{aligned}
\mathrm{Var}\{W(x)\} &= \left\langle p^2(x)/q^2(x)\right\rangle_q - \langle p(x)/q(x)\rangle_q^2 \\
&= \langle W(x)\rangle_p - 1
\end{aligned}$$

Then compute the product of two Gaussian functions and the normalisation constant.

6. **(Estimation of Gaussians)** Consider the following model

$$
\begin{aligned}
\beta &\sim \mathcal{G}(\beta; \nu, 1) \\
\mu &\sim \mathcal{N}(\mu; 0, 1000) \\
x_i &\sim \mathcal{N}(x_i; \mu, \beta^{-1})
\end{aligned}
$$

for $i = 1 \ldots N$. $X = \{x_1, \ldots, x_N\}$. Furthermore, use

$$
\begin{aligned}
\mathcal{G}(\beta; \nu, s) &\equiv \exp((\nu - 1)\log\beta - s\beta - \log\Gamma(\nu) + \nu\log s) \\
\mathcal{N}(x; m, s) &= \exp(-\frac{1}{2s}(x - m)^2 - \frac{1}{2}\log(2\pi s))
\end{aligned}
$$

(a) (4) Derive the full conditionals and give pseudocode of a Gibbs sampler to sample from

$$p(\mu, \beta | X, \nu)$$

(b) (4) Derive the expression for the acceptance probability of a Metropolis-Hastings algorithm to sample from

$$p(\nu | X, \mu, \beta)$$

Use the following proposal $q = \mathcal{G}(\nu; a, 1)$

(c) (4) Combine (a) and (b) and derive a MH algorithm to sample from

$$p(\nu, \mu, \beta | X)$$

(12 points)

The full joint distribution is

$$
\begin{aligned}
\log p(\beta, \mu, x_1, \ldots, x_N | \nu) &= \log\left(\mathcal{G}(\beta; \nu, 1)\mathcal{N}(\mu; 0, 1000)\prod_i^N \mathcal{N}(x_i; \mu, \beta^{-1})\right) \\
&= (\nu - 1)\log\beta - \beta - \log\Gamma(\nu) \\
&\quad -\frac{1}{2000}\mu^2 - \frac{1}{2}\log(2000\pi) \\
&\quad -\frac{\beta}{2}\sum_{i=1}^N x_i^2 - \frac{\beta N}{2}\mu^2 + \beta\mu\sum_{i=1}^N x_i + \frac{N}{2}\log(2\pi\beta)
\end{aligned}
$$

(a) We need the full conditionals

$$
\begin{aligned}
&p(\mu | X, \beta, \nu) \\
&p(\beta | X, \mu, \nu)
\end{aligned}
$$

These are

$$
\begin{aligned}
\log p(\mu | X, \beta, \nu) &=^{+} -\frac{\beta N}{2}\mu^2 + \beta\mu\sum_{i=1}^N x_i - \frac{1}{2000}\mu^2 \\
&=^{+} \log\mathcal{N}(\mu; (\beta N + 1/1000)^{-1}\left(\sum_{i=1}^N x_i\right), (\beta N + 1/1000)^{-1})
\end{aligned}
$$

$$\begin{aligned}
\log p(\beta|X,\mu,\nu) &=^+ (\nu-1)\log\beta - \beta \\
&\quad -\frac{\beta}{2}\sum_{i=1}^{N} x_i^2 - \frac{\beta N}{2}\mu^2 + \beta\mu\sum_{i=1}^{N} x_i + \frac{N}{2}\log(\beta) \\
&= (\nu + N/2 - 1)\log\beta - \left(1 + \frac{1}{2}\sum_{i=1}^{N}(x_i-\mu)^2\right)\beta \\
&=^+ \mathcal{G}(\beta;\nu + N/2, 1 + \frac{1}{2}\sum_{i=1}^{N}(x_i-\mu)^2)
\end{aligned}$$

(b) The target distribution is

$$\log p(\nu|\beta,X,\mu) =^+ (\nu-1)\log\beta - \log\Gamma(\nu)$$

Note that we don't need the normalising constant of the target density for MH. The acceptance probability is simply:

$$\begin{aligned}
\log\alpha(\nu\to\nu') &= \min\{0, \log\frac{\pi(\nu')q(\nu|\nu')}{\pi(\nu)q(\nu'|\nu)}\} \\
\pi(\nu) &= (\nu-1)\log\beta - \log\Gamma(\nu) \\
\log q(\nu|\nu') &= (a-1)\log\nu - \nu - \log\Gamma(a)
\end{aligned}$$

$$\begin{aligned}
\log\alpha &= (\nu'-1)\log\beta - \log\Gamma(\nu') + (a-1)\log\nu - \nu - \log\Gamma(a) \\
&\quad -(\nu-1)\log\beta + \log\Gamma(\nu) - (a-1)\log\nu' + \nu' + \log\Gamma(a) \\
&= (\nu'-\nu)(\log\beta+1) - (\log\Gamma(\nu') - \log\Gamma(\nu)) - (a-1)(\log\nu' - \log\nu)
\end{aligned}$$

(c) Do Gibbs by sampling from the full conditionals. $p(\nu|\cdot)$ will be sampled via MH.

(blank page for further space)

7. (**Metropolis and Gibbs**) x_1 and x_2 are two discrete random variables taking values in $\{-1, 1\}$. Suppose we have the joint distribution $p(x_1, x_2) \propto \exp(\theta x_1 x_2)$.

   Suppose we implement a Metropolis algorithm to sample from this target distribution with the following proposal technique: Given the current configuration $x^{(n)} = (x_1^{(n)}, x_2^{(n)})$, for each n, we choose an index $i^{(n)} \in \{1, 2\}$ randomly with probability 0.5 and flip the sign of $x_{i^{(n)}}$.

   (a) *(2 pts)* Write down the state transition diagram of the proposal distribution and indicate the state transition probabilities,

   (b) *(2 pts)* Find an expression for the acceptance probability as a function of θ,

   (c) *(2 pts)* Write the pseudocode for the Metropolis sampler,

   (d) *(2 pts)* Write down the state transition diagram of the transition Kernel T_M of this Metropolis algorithm and indicate the transition probabilities,

   (e) *(2 pts)* Verify if detailed balance condition is satisfied by this particular Metropolis algorithm for all values of θ.

   (f) *(2 pts)* Write down the pseudocode for a deterministic scan Gibbs sampler.

   (g) *(2 pts)* Write down an expression for the Gibbs transition Kernel T_G in terms of θ.

   (h) *(2 pts)* Verify detailed balance is satisfied the Gibbs transition Kernel T_G for all values of θ.

(16 points)

The target distribution is given by

$\pi(x)$	$x = (-1, -1)$	$x = (1, -1)$	$x = (-1, 1)$	$x = (1, 1)$
	$\exp(\theta)$	$\exp(-\theta)$	$\exp(-\theta)$	$\exp(\theta)$

We define

$$g = \exp(-\theta)/(2\exp(-\theta) + 2\exp(\theta))$$

$\pi(x)$	$x = (-1, -1)$	$x = (1, -1)$	$x = (-1, 1)$	$x = (1, 1)$
	$0.5 - g$	g	g	$0.5 - g$

(a)

The proposal

$q(x'\|x)$	$x = (-1, -1)$	$x = (1, -1)$	$x = (-1, 1)$	$x = (1, 1)$
$x' = (-1, -1)$	0	0.5	0.5	0
$x' = (1, -1)$	0.5	0	0	0.5
$x' = (-1, 1)$	0.5	0	0	0.5
$x' = (1, 1)$	0	0.5	0.5	0

(b) The acceptance probability

$$a(x \to x') = \min\{1, \frac{\pi(x')q(x|x')}{\pi(x)q(x'|x)}\}$$

$a(x \to x')$	$x=(-1,-1)$	$x=(1,-1)$	$x=(-1,1)$	$x=(1,1)$
$x'=(-1,-1)$	1	$(0.5-g)/g$	$(0.5-g)/g$	1
$x'=(1,-1)$	1	1	1	1
$x'=(-1,1)$	1	1	1	1
$x'=(1,1)$	1	$(0.5-g)/g$	$(0.5-g)/g$	1

(c)

$x^{(0)} \sim r(x)$

for $\tau = 1, 2, \ldots$
    $x'_\tau \sim q(x'|x = x^{(\tau-1)})$

    if `rand` $< a(x \to x')$
        $x^{(\tau)} \leftarrow x'_\tau$
    else
        $x^{(\tau)} \leftarrow x^{(\tau-1)}$
    end
endfor

(d)

$$T_M(x'|x) \quad = \quad a(x \to x')q(x'|x) + \delta(x - x') \sum_{x'} (1 - a(x \to x'))q(x'|x)$$

The accept part of the Kernel

$a(x \to x')q(x'\|x)$	$x=(-1,-1)$	$x=(1,-1)$	$x=(-1,1)$	$x=(1,1)$
$x'=(-1,-1)$	0	$(0.5-g)/(2g)$	$(0.5-g)/(2g)$	0
$x'=(1,-1)$	0.5	0	0	0.5
$x'=(-1,1)$	0.5	0	0	0.5
$x'=(1,1)$	0	$(0.5-g)/(2g)$	$(0.5-g)/(2g)$	0

The reject part

$(1-a(x \to x'))q(x'\|x)$	$x=(-1,-1)$	$x=(1,-1)$	$x=(-1,1)$	$x=(1,1)$
$x'=(-1,-1)$	0	$(2g-0.5)/(2g)$	$(2g-0.5)/(2g)$	0
$x'=(1,-1)$	0	0	0	0
$x'=(-1,1)$	0	0	0	0
$x'=(1,1)$	0	$(2g-0.5)/(2g)$	$(2g-0.5)/(2g)$	0

$\delta(x'-x)\sum_{x'}(1-a(x \to x'))q(x'\|x)$	$x=(-1,-1)$	$x=(1,-1)$	$x=(-1,1)$	$x=(1,1)$
$x'=(-1,-1)$	0	0	0	0
$x'=(1,-1)$	0	$(2g-0.5)/g$	0	0
$x'=(-1,1)$	0	0	$(2g-0.5)/g$	0
$x'=(1,1)$	0	0	0	0

$T_M(x'\|x)$	$x=(-1,-1)$	$x=(1,-1)$	$x=(-1,1)$	$x=(1,1)$
$x'=(-1,-1)$	0	$(0.5-g)/(2g)$	$(0.5-g)/(2g)$	0
$x'=(1,-1)$	0.5	$(2g-0.5)/g$	0	0.5
$x'=(-1,1)$	0.5	0	$(2g-0.5)/g$	0.5
$x'=(1,1)$	0	$(0.5-g)/(2g)$	$(0.5-g)/(2g)$	0

(e)

$T_M(x'\|x)\pi(x)$	$x=(-1,-1)$	$x=(1,-1)$	$x=(-1,1)$	$x=(1,1)$
$x'=(-1,-1)$	0	$(0.5-g)/2$	$(0.5-g)/2$	0
$x'=(1,-1)$	$0.5(0.5-g)$	$(2g-0.5)$	0	$0.5(0.5-g)$
$x'=(-1,1)$	$0.5(0.5-g)$	0	$(2g-0.5)$	$0.5(0.5-g)$
$x'=(1,1)$	0	$(0.5-g)/(2)$	$(0.5-g)/(2)$	0

$T_M(x'|x)\pi(x)$ is symmetrical, hence detailed balance is satisfied.

(f)

$x_2^{(0)} \sim q(x_2)$

for $\tau = 1, 2, \ldots$

$x_1^{(\tau)} \sim \pi(x_1|x_2 = x_2^{(\tau-1)})$

$x_2^{(\tau)} \sim \pi(x_2|x_1 = x_1^{(\tau)})$

endfor

(g) Gibbs sampling uses the full conditionals $\pi(x_1|x_2)$ and $\pi(x_2|x_1)$

$\pi(x_1\|x_2)$	$x_2=-1$	$x_2=1$
$x_1=-1$	$1-2g$	$2g$
$x_1=1$	$2g$	$1-2g$

$T_{G,1}(x'\|x)$	$x=(-1,-1)$	$x=(1,-1)$	$x=(-1,1)$	$x=(1,1)$
$x'=(-1,-1)$	$1-2g$	$1-2g$	0	0
$x'=(1,-1)$	$2g$	$2g$	0	0
$x'=(-1,1)$	0	0	$2g$	$2g$
$x'=(1,1)$	0	0	$1-2g$	$1-2g$

$\pi(x_2\|x_1)$	$x_2=-1$	$x_2=1$
$x_1=-1$	$1-2g$	$2g$
$x_1=1$	$2g$	$1-2g$

$T_{G,2}(x'\|x)$	$x=(-1,-1)$	$x=(1,-1)$	$x=(-1,1)$	$x=(1,1)$
$x'=(-1,-1)$	$1-2g$	0	$1-2g$	0
$x'=(1,-1)$	0	$2g$	0	$2g$
$x'=(-1,1)$	$2g$	0	$2g$	0
$x'=(1,1)$	0	$1-2g$	0	$1-2g$

A deterministic scan Gibbs sampler that samples first x_1 then x_2 has the effective transition kernel

$$T_G = T_{G,2} T_{G,1}$$

$T_G(x'\|x)$	$x=(-1,-1)$	$x=(1,-1)$	$x=(-1,1)$	$x=(1,1)$
$x'=(-1,-1)$	$(1-2g)^2$	$(1-2g)^2$	$(1-2g)2g$	$(1-2g)2g$
$x'=(1,-1)$	$(2g)^2$	$(2g)^2$	$(1-2g)2g$	$(1-2g)2g$
$x'=(-1,1)$	$(1-2g)2g$	$(1-2g)2g$	$(2g)^2$	$(2g)^2$
$x'=(1,1)$	$(1-2g)2g$	$(1-2g)2g$	$(1-2g)^2$	$(1-2g)^2$

(h)

It is interesting to note that for the Gibbs sampler $T_G\pi$ **is not symmetric**

$T_G(x'\|x)\pi(x)$	$x=(-1,-1)$	$x=(1,-1)$	$x=(-1,1)$	$x=(1,1)$
$x'=(-1,-1)$	$(1-2g)^2(0.5-g)$	$(1-2g)^2 g$	$(1-2g)2g^2$	$(1-2g)2g(0.5-g)$
$x'=(1,-1)$	$(2g)^2(0.5-g)$	$(2g)^2 g$	$(1-2g)2g^2$	$(1-2g)2g(0.5-g)$
$x'=(-1,1)$	$(1-2g)2g(0.5-g)$	$(1-2g)2g^2$	$(2g)^2 g$	$(2g)^2(0.5-g)$
$x'=(1,1)$	$(1-2g)2g(0.5-g)$	$(1-2g)2g^2$	$(1-2g)^2 g$	$(1-2g)^2(0.5-g)$

Hence the chain is not time reversible, yet, the detailed balance condition holds:

$$\pi(x') = \sum_x T_G(x'|x)\pi(x)$$

The individual factors of the kernel satisfy detailed balance

$T_{G,1}(x'\|x)\pi(x)$	$x=(-1,-1)$	$x=(1,-1)$	$x=(-1,1)$	$x=(1,1)$
$x'=(-1,-1)$	$(1-2g)^2/2$	$(1-2g)g$	0	0
$x'=(1,-1)$	$g(1-2g)$	$2g^2$	0	0
$x'=(-1,1)$	0	0	$2g^2$	$g(1-2g)$
$x'=(1,1)$	0	0	$(1-2g)g$	$(1-2g)^2/2$

$T_{G,2}(x'\|x)\pi(x)$	$x=(-1,-1)$	$x=(1,-1)$	$x=(-1,1)$	$x=(1,1)$
$x'=(-1,-1)$	$(1-2g)^2/2$	0	$(1-2g)g$	0
$x'=(1,-1)$	0	$2g^2$	0	$g(1-2g)$
$x'=(-1,1)$	$g(1-2g)$	0	$2g^2$	0
$x'=(1,1)$	0	$(1-2g)g$	0	$(1-2g)^2/2$