Boğaziçi University, Dept. of Computer Engineering

CMPE 58K, BAYESIAN STATISTICS AND MACHINE LEARNING

Fall 2008, 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 not use any notes, books or calculators.
- Read each question carefully and show all your work. Underline your final answer to each question.
- There are **8** questions. Point values are given in parentheses.
- You have **180 minutes** to do all the problems.

Q	1	2	3	4	5	6	7	8	Total
Score									
Max	12	6	6	6	6	6	6	12	60

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

   (a) *(2 pts)* State the Bayes theorem.

   (b) *(2 pts)* Prove the Bayes theorem.

   (c) *(1 pts)* Which term in the Bayes theorem is called *the prior*.

   (d) *(1 pts)* What is the difference between a state transition diagram and a graphical model?

   (e) *(1 pts)* What is the difference between maximum likelihood and maximum a-posteriori (MAP) estimate of a parameter?

   (f) *(1 pts)* There are two approaches to statistics. One is the Bayesian statistics. What is the other one?

   (g) *(1 pts)* What is a conjugate prior?

   (h) *(1 pts)* What is an undirected graphical model?

   (i) *(1 pts)* What is the marginal likelihood? What is the difference between evidence?

   (j) *(1 pts)* What is an Exponential Family? What is a Canonical Parametrisation?

2. (**could have been a quiz question**) Let x_1 and x_2 are two discrete random variables taking values in $\{-1, 1\}$. We know that $p(x_1 = -1|x_2 = -1) = 1/4$, $p(x_1 = 1|x_2 = 1) = 2/3$, $p(x_2 = -1|x_1 = 1) = 3/7$ and $p(x_2 = 1|x_1 = -1) = 2/3$. Show all your work.

   (a) Find the following quantities

      i. *(2 pts)* Joint: $p(x_1, x_2)$

      ii. *(1 pts)* Marginals: $p(x_1)$, $p(x_2)$

      iii. *(1 pts)* Max-marginal: $\max_{x_1} p(x_1, x_2)$

      iv. *(1 pts)* Covariance of x_1 and x_2

   (b) *(1 pts)* Are x_1 and x_2 independent ? Why or why not?

The given data implies the following conditional distributions:

$p(x_1\|x_2)$	$x_2 = -1$	$x_2 = 1$
$x_1 = -1$	$1/4$	$1/3$
$x_1 = 1$	$3/4$	$2/3$

$p(x_2\|x_1)$	$x_2 = -1$	$x_2 = 1$
$x_1 = -1$	$1/3$	$2/3$
$x_1 = 1$	$3/7$	$4/7$

Since

$$p(x_1|x_2)p(x_2) = p(x_2|x_1)p(x_1)$$

We have

$p(x_1, x_2)$	$x_2 = -1$	$x_2 = 1$
$x_1 = -1$	$p/4$	$(1-p)/3$
$x_1 = 1$	$3p/4$	$2(1-p)/3$

$p(x_2, x_1)$	$x_2 = -1$	$x_2 = 1$
$x_1 = -1$	$q/3$	$2q/3$
$x_1 = 1$	$3(1-q)/7$	$4(1-q)/7$

This implies the (overdetermined) linear system

$$\begin{aligned} p/4 &= q/3 \\ (1-p)/3 &= 2q/3 \\ 3p/4 &= 3(1-q)/7 \\ 2(1-p)/3 &= 4(1-q)/7 \end{aligned}$$

The first two equations imply

$$\begin{aligned} p &= 4q/3 \\ p &= 1-2q \\ 3-6q &= 4q \\ q &= 3/10 \\ p &= 4/10 \end{aligned}$$

The other equations are satisfied

$$\begin{aligned} 3(4/10)/4 &= 3(1-(3/10))/7 \\ 2(1-(4/10))/3 &= 4(1-(3/10))/7 \end{aligned}$$

so the given data is consistent. We find the joint distribution

$p(x_2, x_1)$	$x_2 = -1$	$x_2 = 1$
$x_1 = -1$	1/10	2/10
$x_1 = 1$	3/10	4/10

(6 points)

3. **(Model construction)** Suppose we have a dataset of the exam grades of 200 of students in 3 different subjects: Sports, Maths and History. For each subject we have 3 exam results. Suppose we believe that there are three types of students : Nerdy, Sporty and Arty. We believe that a student can be either nerdy or arty but not both. He or she can be also sporty independent of being nerdy or arty. Given the type of the student, the grade obtained from a subject by a student is assumed to be a random variable. The grade distributions have the same parameters for all students and exams of the same subject. Grade distributions have different parameters for different subjects. Moreover, each student can be ill during an examination, independent of other students and other examinations. If a student is ill during an examination, this would only affect the students performance for that examination.

   (a) *(2 pts)* Carefully define the appropriate random variables to represent this scenario.

   (b) *(2 pts)* Draw the graphical model including model parameters and denote the conditional probability tables

   (c) *(2 pts)* Suppose we wish to find a MAP estimate of the parameters. Write down the loglikelihood function that needs to be optimised with respect to the parameters given the data set. Remember, unknown variables need to be integrated over.

(a)

Indicies

$i \in \{1, \dots, 200\}$	student
$s \in \{\text{Sports, Maths, History}\}$	subject
$e \in \{1, 2, 3\}$	exam

Random variables

$T \in \{\text{Nerdy, Arty}\}$	Type
$S \in \{\text{Sporty, Not Sporty}\}$	Sporty
$I \in \{\text{Ill, not Ill}\}$	Health
G	Grade

Parameters

λ	Parameters

(b)

$$\begin{aligned} G_{i,s,e} &\sim p(G_{i,s,e} | I_{i,s,e}, T_i, S_i, \lambda_s) \\ I_{i,s,e} &\sim p(I_{i,s,e}) \\ T_i &\sim p(T_i) \\ S_i &\sim p(S_i) \\ \lambda_s &\sim p(\lambda_s) \end{aligned}$$

(c)

$$
\begin{aligned}
p(G, I, T, S, \lambda) &= \left(\prod_{i,s,e} p(G_{i,s,e} | I_{i,s,e}, T_i, S_i, \lambda_s) p(I_{i,s,e})\right) \left(\prod_i p(T_i) p(S_i)\right) \left(\prod_s p(\lambda_s)\right) \\
p(G, \lambda) &= \sum_{T_1,\ldots,T_{200}} \sum_{S_1,\ldots,S_{200}} \sum_I p(G, I, T, S, \lambda)
\end{aligned}
$$

(6 points)

4. **(Time Series Modelling)** In the following figures, observations y_t from two processes are given as a function of time index t. Observations are known to be discrete with $y_t \in \{1, \dots, 30\}$. For each realisation, define a plausible process that would generate similar realisations. Define the appropriate latent variables (if you use any), draw the graphical model and provide the conditional probability tables **and/or** state transition diagrams.

   (a) (2 pts)

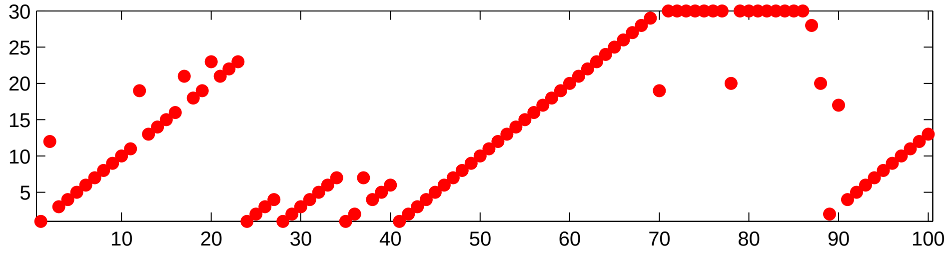


Figure 1: Process 1

   (b) (2 pts)

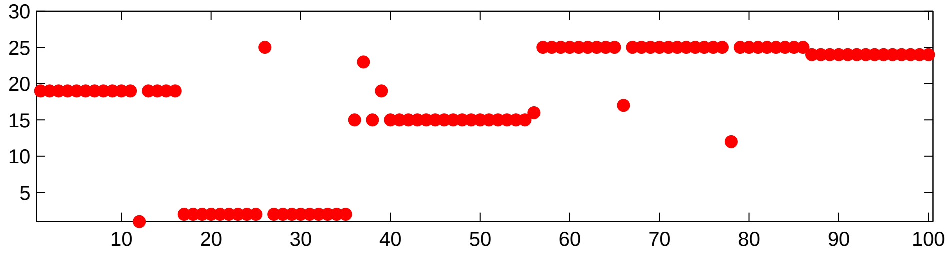


Figure 2: Process 2

   (c) (2 pts)

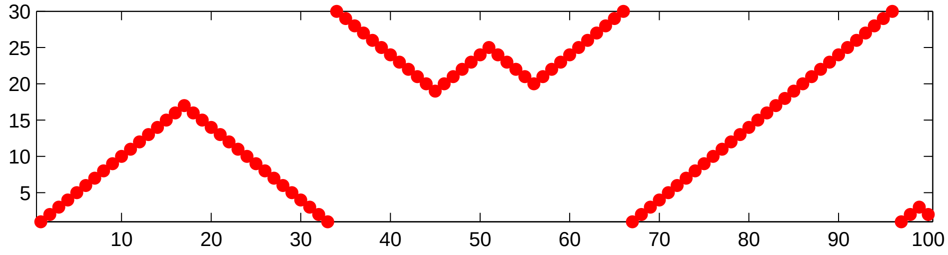
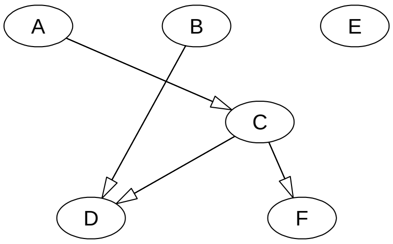


Figure 3: Process 3

5. (**Directed Graphical Models**) Consider the following directed graph G



(a) *(1 pts)* Find a topological ordering of the variables,

(b) *(2 pts)* Write down the implied factorisation of the probability distribution that respects the conditional independence structure implied by G,

(c) *(1 pts)* Draw the associated factor graph,

(d) *(1 pts)* Suppose, each variable has two states. How many free parameters does each conditional probability table have?

(e) *(1 pts)* Draw an equivalent undirected graphical model.

(6 points)

6. (**Partitioned Inverse Equations)** We are given the following partitioned square matrix

$$Z = \begin{pmatrix} A & B \\ D & C \end{pmatrix}$$

We will prove via block Gaussian elimination that

$$Z^{-1} = \begin{pmatrix} A^{-1} + A^{-1}BN^{-1}DA^{-1} & -A^{-1}BN^{-1} \\ -N^{-1}DA^{-1} & N^{-1} \end{pmatrix}$$

where $N = (C - DA^{-1}B)$.

[Hint: First find a factorisation $W = T_1 Z T_2$ where W is block diagonal and T_1 and T_2 are block triangular with identity matrices on the block diagonal. Then argue $Z = T_1^{-1} W T_2^{-1}$ implies $Z^{-1} = T_2 W^{-1} T_1$.]

First thing to check it which Gauss transformation we will need

$$\begin{pmatrix} I & X \\ 0 & I \end{pmatrix}\begin{pmatrix} A & B \\ D & C \end{pmatrix} = \begin{pmatrix} A + XD & B + XC \\ D & C \end{pmatrix}$$

If we wish to render $B + XC = 0$ we need $X = -BC^{-1}$.

$$\begin{pmatrix} I & 0 \\ L & I \end{pmatrix}\begin{pmatrix} A & B \\ D & C \end{pmatrix} = \begin{pmatrix} A & B \\ LA + D & LB + C \end{pmatrix}$$

If we wish to render $LA + D = 0$ we need $L = -DA^{-1}$.

$$\begin{pmatrix} A & B \\ D & C \end{pmatrix}\begin{pmatrix} I & 0 \\ Y & I \end{pmatrix} = \begin{pmatrix} A + BY & B \\ D + CY & C \end{pmatrix}$$

If we wish to render $D + CY = 0$ we need $Y = -C^{-1}D$.

$$\begin{pmatrix} A & B \\ D & C \end{pmatrix}\begin{pmatrix} I & U \\ 0 & I \end{pmatrix} = \begin{pmatrix} A & AU + B \\ D & DU + C \end{pmatrix}$$

If we wish to render $AU + B = 0$ we need $U = -A^{-1}B$.

The 2nd and 4th lead to

$$\begin{aligned} \begin{pmatrix} I & 0 \\ -DA^{-1} & I \end{pmatrix}\begin{pmatrix} A & B \\ D & C \end{pmatrix}\begin{pmatrix} I & -A^{-1}B \\ 0 & I \end{pmatrix} &= \begin{pmatrix} A & B \\ 0 & C - DA^{-1}B \end{pmatrix}\begin{pmatrix} I & -A^{-1}B \\ 0 & I \end{pmatrix} \\ &= \begin{pmatrix} A & 0 \\ 0 & C - DA^{-1}B \end{pmatrix} \end{aligned}$$

If we compute the inverse of both sides

$$\begin{aligned}\begin{pmatrix} A^{-1} & 0 \\ 0 & (C - DA^{-1}B)^{-1} \end{pmatrix} &= \left(\begin{pmatrix} I & 0 \\ -DA^{-1} & I \end{pmatrix} \begin{pmatrix} A & B \\ D & C \end{pmatrix} \begin{pmatrix} I & -A^{-1}B \\ 0 & I \end{pmatrix} \right)^{-1} \\ &= \begin{pmatrix} I & -A^{-1}B \\ 0 & I \end{pmatrix}^{-1} \begin{pmatrix} A & B \\ D & C \end{pmatrix}^{-1} \begin{pmatrix} I & 0 \\ -DA^{-1} & I \end{pmatrix}^{-1}\end{aligned}$$

Which implies

$$\begin{pmatrix} A & B \\ D & C \end{pmatrix}^{-1} = \begin{pmatrix} I & -A^{-1}B \\ 0 & I \end{pmatrix} \begin{pmatrix} A^{-1} & 0 \\ 0 & N^{-1} \end{pmatrix} \begin{pmatrix} I & 0 \\ -DA^{-1} & I \end{pmatrix}$$

(6 points)

7. (**Factorising Gaussians**)

Given the model

$$\begin{aligned} x_0 &\sim \mathcal{N}(x_0; 0, \Sigma) \\ x_1 | x_0 &\sim \mathcal{N}(x_1; A x_0, Q) \end{aligned}$$

where

$$\mathcal{N}(x; \mu, \Sigma) \equiv |2\pi\Sigma|^{-1/2} \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right)$$

is the multivariate Gaussian distribution, A, Q and Σ are known matrices.

(a) *(2 pts)* Find the joint distribution $p(x_0, x_1)$ and express it as a multivariate Gaussian.

(b) *(4 pts)* Find a factorisation of $p(x_0, x_1)$ as $p(x_1)p(x_0|x_1)$ and express the factors as Gaussian distributions.

(a) First solution, via moments

$$\begin{aligned} x_0 &\sim \mathcal{N}(x_0; 0, \Sigma) \\ \epsilon &\sim \mathcal{N}(\epsilon; 0, Q) \\ x_1 &= A x_0 + \epsilon \\ \langle x_1 | x_0 \rangle &= A x_0 \\ \langle x_1 \rangle &= \langle A x_0 \rangle = 0 \\ x_1 &= A x_0 + \epsilon \\ x_1 x_1^\top &= (A x_0 + \epsilon)(A x_0 + \epsilon)^\top = A x_0 x_0 A^\top + 2\epsilon x_0^\top A^\top + \epsilon\epsilon^\top \\ \left\langle x_1 x_1^\top \right\rangle &= A \langle x_0 x_0 \rangle A^\top + \langle \epsilon \rangle \left\langle x_0^\top \right\rangle A^\top + A \langle x_0 \rangle \left\langle \epsilon^\top \right\rangle + \left\langle \epsilon\epsilon^\top \right\rangle = A \langle x_0 x_0 \rangle A^\top + \left\langle \epsilon\epsilon^\top \right\rangle \\ &= A\Sigma A^\top + Q \\ \mathrm{Cov}\left[x_1\right] &= \left\langle x_1 x_1^\top \right\rangle - \langle x_1 \rangle \left\langle x_1^\top \right\rangle = \left\langle x_1 x_1^\top \right\rangle = A\Sigma A^\top + Q \\ \left\langle x_0 x_1^\top \right\rangle &= \left\langle x_0 x_0^\top A^\top + \epsilon^\top \right\rangle = \left\langle x_0 x_0^\top \right\rangle A^\top = \Sigma A^\top \end{aligned}$$

$$p(x_0, x_1) \propto \exp\left(-\frac{1}{2}\begin{pmatrix} x_0 \\ x_1 \end{pmatrix}^\top \begin{pmatrix} \Sigma & \Sigma A^\top \\ A\Sigma & A\Sigma A^\top + Q \end{pmatrix}^{-1} \begin{pmatrix} x_0 \\ x_1 \end{pmatrix}\right)$$

Second solution, via canonical parameters

$$\begin{aligned} p(x_0) &= \mathcal{N}(x_0; 0, \Sigma) \propto \exp\left(-\frac{1}{2} x_0^\top \Sigma^{-1} x_0\right) \\ p(x_1|x_0) &= \mathcal{N}(x_1; A x_0, Q) \propto \exp\left(-\frac{1}{2} x_1^\top Q^{-1} x_1 + x_0^\top A^\top Q^{-1} x_1 - \frac{1}{2} x_0^\top A^\top Q^{-1} A x_0^\top\right) \\ p(x_0, x_1) &\propto \exp\left(-\frac{1}{2}\begin{pmatrix} x_0 \\ x_1 \end{pmatrix}^\top \begin{pmatrix} \Sigma^{-1} + A^\top Q^{-1} A & -A^\top Q^{-1} \\ -Q^{-1} A & Q^{-1} \end{pmatrix} \begin{pmatrix} x_0 \\ x_1 \end{pmatrix}\right) \end{aligned}$$

To find the covariance, we evoke the block matrix inverse equations, this leads to the following blocks

$$\begin{pmatrix} \Sigma^{-1} + A^\top Q^{-1} A & -A^\top Q^{-1} \\ -Q^{-1} A & Q^{-1} \end{pmatrix}^{-1} = \begin{pmatrix} \Lambda_1 & \Lambda_{12} \\ \Lambda_{12}^\top & \Lambda_2 \end{pmatrix}$$

$$\begin{aligned}
\Lambda_1 &= \left((\Sigma^{-1} + A^\top Q^{-1} A) - (-A^\top Q^{-1})(Q^{-1})^{-1}(-Q^{-1}A)\right)^{-1} = \Sigma \\
\Lambda_{12} &= -\left((\Sigma^{-1} + A^\top Q^{-1} A) - (-A^\top Q^{-1})(Q^{-1})^{-1}(-Q^{-1}A)\right)^{-1}(-A^\top Q^{-1})(Q^{-1})^{-1} = \Sigma A^\top \\
\Lambda_2 &= (Q^{-1})^{-1} + (Q^{-1})^{-1}(-Q^{-1}A)\left((\Sigma^{-1} + A^\top Q^{-1} A) - (-A^\top Q^{-1})(Q^{-1})^{-1}(-Q^{-1}A)\right)^{-1}(-A^\top Q^{-1})(Q^{-1})^{-1} \\
&= Q + A\Sigma A^\top
\end{aligned}$$

(b)

We just evoke the block matrix inverse

$$\begin{pmatrix} \Sigma & \Sigma A^\top \\ A\Sigma & A\Sigma A^\top + Q \end{pmatrix}^{-1} = \begin{pmatrix} \Sigma^{-1} + A^\top Q^{-1} A & -A^\top Q^{-1} \\ -Q^{-1} A & Q^{-1} \end{pmatrix}$$

$$\begin{aligned}
p(x_0, x_1) &\propto \exp\left(-\frac{1}{2}x_0^\top \Sigma^{-1} x_0 - \frac{1}{2}x_1^\top Q^{-1} x_1 + x_0^\top A^\top Q^{-1} x_1 - \frac{1}{2}x_0^\top A^\top Q^{-1} A x_0^\top\right) \\
p(x_0|x_1) &\propto \exp\left(-\frac{1}{2}x_0^\top (\Sigma^{-1} + A^\top Q^{-1} A) x_0 + x_0^\top A^\top Q^{-1} x_1\right) \\
&\propto \mathcal{N}(x_0; (\Sigma^{-1} + A^\top Q^{-1} A)^{-1} A^\top Q^{-1} x_1, (\Sigma^{-1} + A^\top Q^{-1} A)^{-1}) \\
p(x_1) &= \mathcal{N}(x_1; 0, A\Sigma A^\top + Q)
\end{aligned}$$

(6 points)

(blank page for further space)

8. (**Metropolis and Gibbs**) x_1 and x_2 are two discrete random variables taking values in $\{-1, 1\}$. Suppose we have the joint distribution $p(x_1 = a, x_2 = b) = \pi_{a,b}$. We further have $g = \pi_{-1,1} = \pi_{1,-1} > \pi_{1,1} = \pi_{-1,-1}$.

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

   (b) *(1 pts)* Find an expression for the acceptance probability as a function of g,

   (c) *(1 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 (i.e., if $T_M(x|x')\pi(x') = T_M(x'|x)\pi(x)$) for all values of g.

   (f) *(1 pts)* Suppose we also implement a deterministic scan Gibbs sampler (that is we sample alternatingly from the full conditional distributions $p(x_1|x_2)$ and $p(x_2|x_1)$). Write down the pseudocode.

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

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

The target distribution is given by

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

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

$a(x \to x')$	$x = (-1, -1)$	$x = (1, -1)$	$x = (-1, 1)$	$x = (1, 1)$
$x' = (-1, -1)$	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 $\tau = 1, 2, \dots$
    $x_1^{(\tau)} \sim \pi(x_1|x_2 = x_2^{(\tau-1)})$

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

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

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

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

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

(12 points)