EECS

KTH

Probabilistic Graphical Models DD2420

Exam 8:00-13:00 June 4, 2020

**Aids:** The Course Book, calculator/math app

**Help: No other help is allowed!**

**Upload a scan or photo of each answer (pdf, jpg, png) by 13:00.**

No answers accepted after CANVAS has closed the assignment.

Time, for zoom oral examination on Friday for those passing the written part, will be emailed on Thursday. That oral examination is required for passing the exam.

**Observe:**

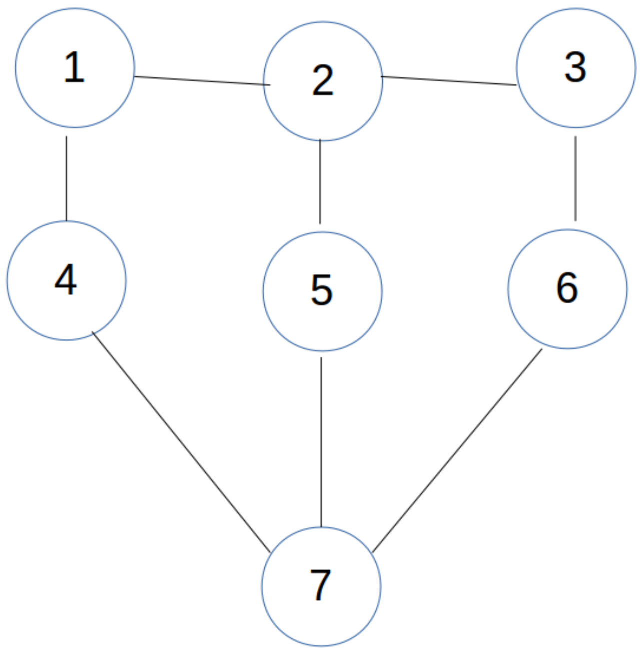
- Name and person number on every page
- Answers should be in English or Swedish
- Only write on one side of the sheets
- Specify the total number of handed in pages on the cover
- Be careful to label each answer with the question number and letter
- All questions should be answered briefly but do motivate your answer and clearly state any additional assumptions you may need to make.

**Responsible:** John Folkesson, 08-790-6201

**Part A (40 points - passing is 50%) Approximate Inference**

1. Loopy Belief Propagation (10 points):

   The parts here will refer to this MRF:




   The nodes are all associated with binomial random variables $X_i$ with $i \in \{1, 2, 3, 4, 5, 6, 7\}$ and $X_i \in \{-1, 1\}$. The edge potentials are $\phi_{ij} = 1 + X_i * X_j$.

   a) Draw the cluster graph for the above undirected network (MRF) where each maximal clique in the graph forms a cluster. Indicate the sepsets on each edge. (1p)

   b) Set up Loopy belief propagation and calibrate using the sum product (SP) belief propagation on the cluster graph. What would a SP-message between two nodes look like? What is $P(X_1 = 1, X_2 = -1)$=? What is $P(X_1 = 1, X_2 = 1)$=? (6p).

   c) Add edges to the graph to triangulate it and draw the resulting cluster graph. Label the sepsets of the edges of the new graph in a way that satisfies the loopy running intersection property (3p).

2. Variational methods (10 points):

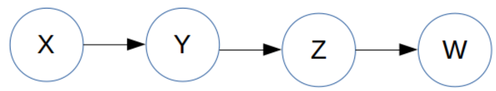
   a) $p(\mathbf{Z} \mid \mathbf{X} = \mathbf{x})$ is a distribution with $\mathbf{x}$ being some observed data and $\mathbf{Z}$ a vector of hidden (latent) random variables. The joint distribution is $p(\mathbf{Z}, \mathbf{X})$ and the prior on $\mathbf{X}$ is $p(\mathbf{X})$. $q(\mathbf{Z})$ is our variational approximation. Explain fully how the data shapes our $q(\mathbf{Z})$ when using variational methods. Include the Kulback-Liebler divergence and the ELBO. (5p)

   b) $P(X = 0) = .2,$, $P(X = 1) = .4,$, $P(X = 2) = .1,$, and $P(X = 3) = .3,$. What is the entropy of $P(X)$? (3p)

   c) $p(X)$ is a pdf with expectation value for $X$ equal to 1 and the expectation value for $X^2$ equal to 5. We wish to approximate this with the M-Projection to a Gaussian distribution. Explain what an M-projection is. What is that Gaussian distribution? (2p)

3. Sampling based Approximation (10 points):

Use this Bayesian Net model for the following questions:




| | x=0 | x=1 |
|---|---|---|
| P(x) | .31 | .69 |

| $P(y \mid x)$ | y=0 | y=1 |
|---|---|---|
| x=0 | .32 | .68 |
| x=1 | .55 | .45 |

| $P(z \mid y)$ | z=0 | z=1 |
|---|---|---|
| y=0 | .85 | .15 |
| y=1 | .25 | .75 |

| $P(w \mid z)$ | w=0 | w=1 |
|---|---|---|
| z=0 | .99 | .01 |
| z=1 | .33 | .67 |

a) Draw 5 samples of $(x, y)$ given the following numbers drawn uniformly from the interval (0,1):

for x: $\{0.9, 0.6, 0.1, 0.4, 0.3\}$

for y: $\{0.2, 0.8, 0.5, 0.7, 0.9\}$ (3p)

b) We have evidence that $z = 1$. Compute the normalized weights of each of your samples. Resample the set from (a) using these 5 uniform random numbers on (0,1): $\{0.2, 0.5, 0.8, 0.1, 0.7\}$. Be clear on what samples from (a) are being resampled and why. (4p)

c) We now get a measurement of w=1. Use rejection sampling, your samples from (b) and these 5 uniform random numbers on (0,1): $\{0.8, 0.4, 0.2, 0.9, 0.3\}$ to sample from the distribution with this additional evidence. (3p)

4. Markov Chain Monte Carlo Methods, MCMC (10 points)

Consider the distribution

$p(x) \propto (\frac{sin(x)}{x})^2.$

with x in radians. (Recall that this will be 1 in the limit as x goes to zero.) And the kernel

$T(x \rightarrow x') = \frac{1}{2\pi}$ for $x' \in (x - \pi, x + \pi)$.

Start the chain at $x_0 = 0$ and use the Metropolis-Hastings algorithm with the following samples from a uniform distribution on (0,1):

$\{0.5, 0.25, 0.75\}$

to generate the next three samples from the chain. To test acceptance use the following three samples from a uniform distribution on (0,1):

$\{0.05, 0.1, 0.3\}$

**Part B (20 points - passing is 50%) Learning**

5. Estimation (10 points)

   Our model of a process that generates $x$ values has 2 parameters, a and b, and looks like this:

   $p(x) = \frac{1}{2b}$ for $x \in (a - b, a + b)$; $a \in (-\infty, \infty)$ and $b \in [0, \infty)$.

   the prior on these parameters is:

   $a \sim N(0, 4)$, the Gaussian distribution with mean 0 and variance 4.

   $p(b) = e^{1-b}$ for $b \geq 1$ and $p(b) = 0$ for $b < 1$.

   If we observe the following values generated by the process:

   $x_0 = 1$, $x_1 = 1.2$, $x_2 = 0.5$ and $x_3 = -0.2$.

   What are the maximum likelihood estimate, MLE, and maximum a posteori, MAP, estimate for a and b?

6. Partially Observed Data (10 points)

a) We want to determine the proportion of people with corvid-19. We test people at random by stopping them on the street. We realize that people staying at home including people with symptoms will not be part of our study and thus can be considered missing data. We also observe this variable 'staying at home'. Discuss this in terms of missing at random and missing completely at random. What can we say from our results? What will we not be able to say? Draw a Bayesian Network that describes the study. (4p)

b) We are studying the effects of nutrition on speed of mice through a maze. We have two sets of mice given different diets. We select one mouse from each group and let them race. We record the speed for each, $s_1$ and $s_2$. We measure what we think is a significant difference in speed. Unfortunately, we can not tell the mice apart so we do not know which is from which group. What name do we give to this situation? What conclusions may we infer? Draw a Bayes Net for this situation.(3p)

c) We have a model of a process with parameters $\mathbf{y}$. We measure some data $\mathbf{x}$, that should depend on the parameters, over three days. Afterwards we notice we ran out of disk space and lost all the data from day 3. If we are now going to use the data to estimate the parameters what problems might there be in terms of MAR and MCAR? What conclusions may we infer? Draw a Bayes Net for this. (3p)