Problem Sheet 3

CMPE 58K, Bayesian Statistics and Machine Learning

Instructor: A. T. Cemgil

Due: 5 Nov 2008, 10:00.

Exercises are labelled with Greek characters α, β, γ and π. Each label denotes the type of the question and roughly corresponds to its difficulty with α the hardest. I **don't** expect you to solve all the questions but you should solve **at least one new** question of each type. A π denotes questions that have a programming component; these are **no longer optional**, i.e., you have solve at least one of these. Don't send any executables or mfiles, just printout the source code and a few example run outputs. However, write your programs clearly as some of these may be used as subroutines in later exercises.

Note that handing in four questions is the minimum requirement; you can always solve and submit more – needless to say the more, the better.

You can hand in new questions from Assignment sheet 2, and these count towards the minimum requirement, however you must hand in at least one question from Sheet 3.

A3.1 (γ) (**Sequential application of the Bayes Theorem**) Recall problem $A1.3$ from assignment sheet 1, where we have three coloured boxes r (red), b (blue), and g (green). Box r contains 3 apples, 4 oranges, and 3 limes, box b contains 1 apple, 1 orange, and 0 limes, and box g contains 3 apples, 3 oranges, and 4 limes. Boxes are chosen in sequence according to the following rules:

- If $t = 0$, choose a box with probabilities $p(r) = 0.2$, $p(b) = 0.2$, $p(g) = 0.6$.
- If t is odd, choose another box with equal probability, that is different from the current box.
- If t is even, choose another box with equal probability, that is different from the current box and choose a fruit with replacement.

(a) Choose the appropriate random variables, write down the generative model and draw the associated directed graphical model.

(b) Draw a state transition diagram (for the boxes only).

(c) Define the conditional probability tables given the rules above.

(d) Write a program to find numerically the probability of selecting the red (blue, gree) box at a given t. Plot the probabilities as a function of t for $t = 1, 3, \ldots, 50$.

(e) If we observe that the first selected fruit is a lime and the second fruit is an orange, what is the probability that the current box is red (blue, green)?

A3.2 (β) (**Beta function**) In this exercise, we prove that the beta distribution, given by

$$\mathcal{B}(w; a, b) \equiv \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} w^{a-1}(1-w)^{b-1}$$

is correctly normalized. This is equivalent to showing that

$$\int_0^1 w^{a-1}(1-w)^{b-1} dw = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \tag{1}$$

(a) Show that (1) is true.
[Hint: Consider the hint in Bishop, problem 2.5, pp128]

(b) Using (1), show that

$$\begin{aligned} \langle w \rangle &= \frac{a}{a+b} \\ \left\langle w^2 \right\rangle - \langle w \rangle^2 &= \frac{ab}{(a+b)^2(a+b+1)} \\ w^* &= \arg\max_w \mathcal{B}(w; a, b) = \frac{a-1}{a+b-2} \qquad a, b > 1 \end{aligned}$$

A3.3 (γ) (**Differential Entropy of a Gaussian**) Find

$$H[q] = -\langle \log q(x) \rangle_q$$

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

A3.4 (β) (**Differential Entropy of a Beta distribution**) Find

$$H[q] = -\langle \log q(x) \rangle_q$$

where $q(x) = \mathcal{B}(a, b)$.

A3.5 (β) (**Inverting the arrow in a Gaussian network**) Given a factorisation of the form $p(y|x)p(x)$ where

$$\begin{aligned} x &\sim \mathcal{N}(x; \mu, \Sigma) \\ y|x &\sim \mathcal{N}(y; Cx, R) \end{aligned}$$

Express this distribution in form of $p(x|y)p(y)$.

A3.6 (β) (**The nasty lecturer**) Every week k, a class of students have to write a quiz, if the random variable $r_k = 1$. The model for the quizes is as follows:

$$\pi | a \sim \mathcal{B}(a, 2) \tag{2}$$
$$r_k | \pi \sim \mathcal{BE}(r_k; \pi) \tag{3}$$

Here, the $\mathcal{B}$ and $\mathcal{BE}$ are Beta and Bernoulli distributions respectively. Suppose, we have observed the values of $r_1, r_2, \ldots, r_n$. We let $r_{1:k}$ detote $r_1, r_2, \ldots, r_k$.

(a) Draw the directed graphical model for the generative model. Include parameter a also as a random variable.

(b) Suppose $a = \sqrt{3}/2$. Compute the probability that there will be a quiz at week $k = n + 1$. Draw also the factor graph for this problem.

(c) Suppose a is unknown. Find the log-likelihood function for a, $\log p(r_{1:k}|a)$.

(d) Assume that $p(a)$ is uniform on $[0.1, 5]$. Write a program to compute and plot the posterior density $p(a|r_{1:k})$ numerically. Make sure that the density is normalised. Plot the posterior densities $p(a|r_{1:k})$ for $k = 1 \dots 5$ for the following observation sequence $r = [10011]$. For example for $k = 2$, you plot $p(a|r_1 = 1, r_2 = 0)$ and ignore r_3, r_4 and r_5. For each k, compute the mean and variance of the posterior distribution.

(e) Consider $p(r_{k+1}|r_{1:k}, a)$ and $p(r_{k+1}|r_{1:k})$. Are those quantities different from each other?

A3.7 (β) (**The nastier lecturer**) Repeat question for the model in question A3.6

$$\begin{aligned} \pi_k|a &\sim \mathcal{B}(a, 2) \\ r_k|\pi_k &\sim \mathcal{BE}(r_k; \pi_k) \end{aligned}$$

and comment, how this model is different from the one given in Eq.(2) and Eq.(3).

A3.8 (α^*, π) (**Self Localisation**) A Robot is moving across a circular coridor. We assume that the possible positions of the robot is a discrete set with N locations. The initial position of the robot is unknown and assumed to be uniformly distributed. At each step k, the robot stays where it is with probability ϵ, or moves to the next point in counterclock direction with probability $1 - \epsilon$. At each step k, the robot can observe the true location with probability w. With probability $w - 1$, the location sensor fails and gives a measurement that is independent from the true position.

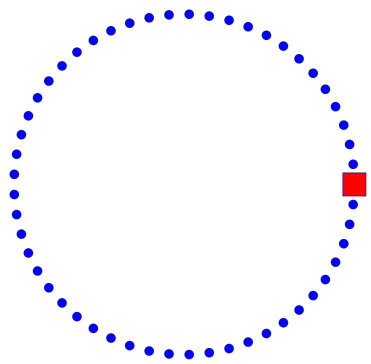

Figure 1: Robot (Square) moving in a circular corridor. Small circles denote the possible N locations.

(a) Choose the appropriate random variables, define their domains, write down the generative model and draw the associated directed graphical model.

(b) Define the conditional probability tables given the verbal description above.

(c) Specify the following verbal statements in terms of posterior quantities using mathematical notation.

[Hint: for example "the distribution of the robots location two time step later given its current position at time k" should be answered as $p(s_{k+2}|s_k)$]

- Distribution of the robots current position given the observations so far,
- Distribution of the robots next position given the observations so far,
- Distribution of the robots next sensor reading given the observations so far,
- Distribution of the robots initial position given observations so far,
- Marginal Distributions of the robots positions at the past given observations so far,
- Most likely current position of the robot given the observations so far,
- Most likely trajectory taken by the robot from the start until now given the observations so far,

(d) Implement a program that simulates this scenario; i.e., generates realisations from the movements of the robot and the associated sensor readings.

[Hint: You can use the `randgen` function you wrote earlier. Simulate a scenario for $k = 1 \dots 100$ with $N = 50, \epsilon = 0.3, w = 0.8$]

(e) (Optional) Implement a program that computes the posterior quantities in 8c, given the sensor readings.

(f) (**The kidnap**) Assume now that at each step the robot can be kidnapped with probability κ. If the robot is kidnapped its new position is independent from its previous position and is uniformly distributed. Repeat 8d and 8e for this new model with $\kappa = 0.1$.

[Hint: Can you reuse your code?]

A3.9 (α, π) (**Self Localisation on prime numbers**) A robot is moving across a circular coridor. We assume that the possible positions of the robot are elements of a discrete set with N locations, numbered as $i = 1, \dots, N$. The exact initial position of the robot is unknown but it is known to be located on one of the non-prime locations. At each step k, the robot stays where it is with probability ϵ, or moves to the next point in counterclock direction with probability $1 - \epsilon$.

At each step k, the robot can observe the color of the tile it is on, independently from previous readings. The corridor is designed such that the tiles with prime numbered locations $i = 2, 3, 5, 7, 11, \dots$ are white, others are blue. Due to the noise present at the visual sensor, with probability δ a white (blue) tile is observed as blue (white).

(a) Define the conditional probability tables given the verbal description above. Write a program that generates one given N, δ and *epsilon*.

[Hint: Matlab has a function called `isprime`.]

(b) Implement a program that simulates this scenario; i.e., generates realisations from the movements of the robot and the associated sensor readings.

[Hint: You can use the `randgen` function you wrote earlier. Simulate a scenario for $k = 1, 2, \dots K$ with sufficiently large K for $N = 50, \epsilon = 0.3, \delta = 0.9$]

(c) (Optional) Implement a program that computes, for each time step, the posterior distribution over the robots position, given the sensor readings so far.

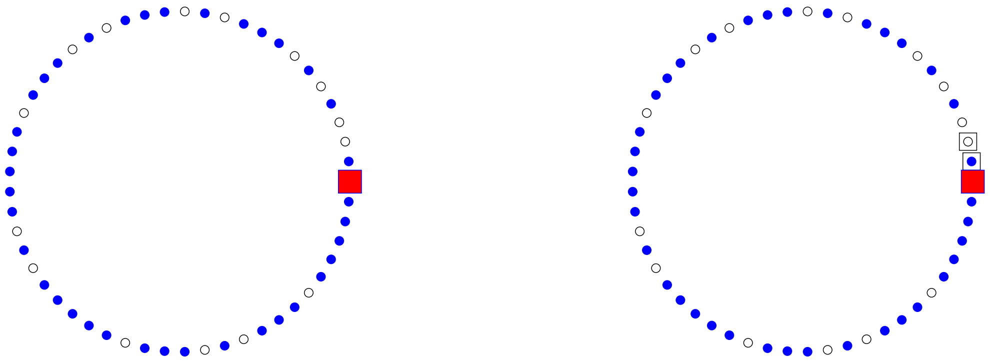

Figure 2: (Left) Robot (Square) moving in a circular corridor. Small circles denote the possible N tiles. Prime numbered tiles are white, others are blue. The robot can sense the color of the tile it is on. (Right) The sensor can sense the color of two tiles in front.

(d) Suppose we modify the sensor such that it can sense the color of two tiles in front (in counterclock direction), independent from each other and independent from previous readings. Define the appropriate random variables.