Probabilistic Models: Spring 2013
Exercise session 2 (exercises 5–9)

Instructions: All course participants are requested to submit their exercise solutions as follows:

- Deadline: before 12 o'clock noon of the Wednesday when the corresponding exercise session will be held
- Submission as a PDF file by email to joonas.paalasmaa at helsinki.fi (cc: to petri.myllymaki at cs.helsinki.fi)
- If you write the solutions by hand, please scan your paper. However, we strongly recommend that you type your solution by using a word processor. LaTeX is of course especially suitable for typesetting math, and it also has a convenient front-end LyX.
- Use as the title of your paper:
  "ProMo-2013, Exercise session n, yourlastname."
- Use as the subject line of the message:
  "ProMo-2013, Exercise session n, yourlastname."
- In all the exercises, do not just give the answer, but also the derivation of how you obtained it.
- Participants are encouraged to write computer programs to derive solutions whenever appropriate. In this case, please enclose the program source code too as a separate file.

After the exercise session, you are allowed to send a modified version of ONE of the solutions you sent before the exercise session:

- Deadline: midnight after the exercise session.
- Submission as before, but this time please send only the modified solution, not the other (unchanged) solutions.
- Enclose the original solution first and then continue with the new material: first explain what you did wrong in the first time and then continue with modifications.
- As the title of the submission (and subject line for the email message), please use:
  "ProMo-2013, Exercise session n, yourlastname, modified exercise x."

5. Let us consider a coin tossing experiment as a Bernoulli model. Let us assume the probability of getting heads in a trial, i.e., $P(H) = \theta$. (T denotes getting tails).

Now assume the outcome of the above experiment is $D = \{HHTHTTHHTHTTHHH\}$.

a) Calculate the maximum likelihood parameters.

b) Also calculate the likelihood of the data with these parameters.

c) Calculate the likelihood when the coin is fair.

d) Calculate posterior distribution, i.e., $P(\theta \mid D)$ considering uniform and Jeffreys priors.

e) Now in the given order, for each observation X, calculate its predictive probability given the preceding sequence, i.e., (in this case) calculate $P(X = H)$, $P(X = H \mid D = \{H\})$, $P(X = T \mid D = \{HH\})$, $P(X = H \mid D = \{HHT\})$ and so on (all 15 probabilities) using maximum likelihood parameters and Bayesian inference (all parameters) with the uniform prior. Also calculate the product of these 15 predictive probabilities.

6. Do the above calculations when the outcome of the experiment is $D = \{TTTHHHHHHH\}$.

7. Let us consider a 6-sided dice rolling experiment as a multinomial model (i.i.d. multi-valued Bernoulli). We roll the dice 50 times, and observe data D with the following counts for the sides:

$$X = 1 : 8 \text{ times}$$
$$X = 2 : 4 \text{ times}$$
$$X = 3 : 9 \text{ times}$$
$$X = 4 : 7 \text{ times}$$
$$X = 5 : 12 \text{ times}$$
$$X = 6 : 10 \text{ times}$$

(a) Calculate the maximum likelihood parameters, given the above data.

(b) Calculate the posterior distribution, i.e., $P(\theta_1, \theta_2, \ldots, \theta_6 \mid D)$ considering (i) the uniform prior $\text{Dir}(1, 1, 1, 1, 1, 1)$ and (ii) the Jeffreys prior $\text{Dir}(0.5, 0.5, 0.5, 0.5, 0.5, 0.5)$.

(c) Using Bayesian inference with the uniform prior, calculate the predictive distribution (all 6 probabilities) of the next result given D.

(d) Give an example of a Dirichlet prior distribution so that $P(X = 3 \mid D) = P(X = 5 \mid D) = 0.25$.

8. Let $P(D \mid \theta)$ be a Bernoulli likelihood and $P(\theta)$ a beta prior with hyper-parameters α and β. The data D is collected by drawing balls with replacement and the total N observations are divided into N_b black balls and N_w white ones. Prove that the maximum a posteriori estimate is

$$\theta_{MAP} = \arg\max_{\theta} P(D \mid \theta)P(\theta) = \frac{\alpha + N_b - 1}{\alpha + \beta + N_b + N_w - 2}.$$

9. Medical diagnosis. Let's have the following notation:

Notation	Explanation
$A = 1$	A person has brain cancer
$B = 1$	A person has a high blood calcium level
$C = 1$	A person has a brain tumor
$D = 1$	A person has seizures that cause unconsciousness
$E = 1$	A person has severe headaches

An expert have told us the following information about the relationships between variables:

> Probability of severe headaches $P(E = 1)$ depends only on the fact whether a person has a brain tumor (C) or not. On the other hand, if one knows the blood calcium level (B) and whether the person has a tumor or not (C), one can specify the probability of unconsciousness seizures $P(D = 1)$. In this case, the probability of D doesn't depend on the presence of the headaches (E) or (directly) on the fact whether the person has brain cancer or not (A). The probability of a brain tumor (C) depends directly only on the fact, whether the person has brain cancer or not (A).

Construct a DAG that represents (exactly) the conditional independencies specified by the expert.