## 8 Machine Learning and Bayesian Inference (sbh11)

You are playing a game against an opponent who has a biased die with probabilities $\{d_1, \ldots, d_6\}$. For each outcome of the die there is a biased coin. The coins show a head with probabilities $p_i$ for $i = 1, \ldots, 6$. Your opponent produces a sequence $\mathbf{o} = (o_1, \ldots, o_m)$ of heads (H) and tails (T) having length $m$. Each is generated by first rolling the die, then flipping the coin corresponding to the die's outcome. Let the random variable (RV) denoting the $i$th outcome be $O_i \in \{H, T\}$, and the RV denoting the outcome of the $i$th roll of the die be $D_i \in \{1, \ldots, 6\}$.

(*a*) Write down an expression for $p_{i,j} = \Pr(O_i | D_i = j)$. [2 marks]

(*b*) Collecting the parameters describing the die and the coins into a vector $\boldsymbol{\theta}$, show that the log-likelihood for the observed outcomes is

$$\log \Pr(\mathbf{o}|\boldsymbol{\theta}) = \sum_{i=1}^{m} \log \sum_{j} p_{i,j} d_j.$$

[4 marks]

(*c*) Define the *latent variables* $z_i^{(j)}$ taking value 1 if the $i$th outcome is generated by the $j$th coin and 0 otherwise. Show that

$$\log \Pr(\mathbf{o}, \mathbf{Z}|\boldsymbol{\theta}) = \sum_{i} \sum_{j} z_i^{(j)} (\log p_{i,j} + \log d_j)$$

where $\mathbf{Z}$ collects together all the values for the latent variables. [4 marks]

(*d*) The *Expectation Maximization (EM) Algorithm* defines the expression

$$L(q, \boldsymbol{\theta}) = \sum_{\mathbf{Z}} q(\mathbf{Z}) \log \frac{\Pr(\mathbf{o}, \mathbf{Z}|\boldsymbol{\theta})}{q(\mathbf{Z})}$$

for an arbitrary distribution $q$, and relies on the fact that

$$L(q, \boldsymbol{\theta}) = \log \Pr(\mathbf{o}|\boldsymbol{\theta}) - D_{\mathrm{KL}}(q(\mathbf{Z}) || \Pr(\mathbf{Z}|\mathbf{o}, \boldsymbol{\theta}))$$

where $D_{\mathrm{KL}}(.||.)$ denotes Kullback-Liebler distance. Explain how these expressions lead to the two steps used by the EM algorithm to maximize the likelihood $\log \Pr(\mathbf{o}|\boldsymbol{\theta})$. [5 marks]

(*e*) Derive the *E step* of the EM algorithm for maximizing the likelihood in the case of the die and coins problem. Your answer should include an expression for the resulting probability distribution in terms of the parameters $\boldsymbol{\theta}$. [5 marks]