BCCS 2008/09: Graphical models and complex stochastic systems: Exercises 3

1. Suppose that for three random variables x, y, z, the joint distribution factorises as a product of a function of x and z, and a function of y and z, i.e. $p(x, y, z) = f(x, z)g(y, z)$ (where f and g are arbitrary functions, not assumed to be p.d.f.'s). Prove that $x \perp\!\!\!\perp y \mid z$, assuming that the random variables are discrete.

2. The following table gives the (fictitious) admission rates for different departments of a university by sex.

Dept.	Sex	no. applying	no. admitted
I	Male	100	25
	Female	300	75
II	Male	200	100
	Female	200	100
III	Male	300	225
	Female	100	75

   Let X be the binary variable for sex, Y the binary variable indicating admission and Z the variable indicating the department. Investigate whether $X \perp\!\!\!\perp Y$ andor $X \perp\!\!\!\perp Y \mid Z$. Explain your results.

3. Consider a variant of the '10+1' coin tossing problem from the 1st lecture, where instead of a discrete choice between 2 biased coins, the parameter θ is supposed to be drawn from a Beta(α, β) prior: $p(\theta) = [\Gamma(\alpha + \beta)/\Gamma(\alpha)\Gamma(\beta)]\theta^{\alpha-1}(1 - \theta)^{\beta-1}$. Write down the joint distribution $p(\theta, x, y)$. Integrate out θ to find $p(x, y)$. Indicate how you would use this to find the conditional expectation $E(y|x)$. Note that you get the answer much more easily by first finding the posterior $p(\theta|x)$, and then noting that $E(y|x) = P\{y = 1|x\} = E(\theta|x)$ (which we already know, or can easily find).

4. Consider the following possible factorisations for the joint distributions of all the variables mentioned. For each, if possible, draw the corresponding DAG. If not possible, say why.

   i. $p(a)p(b|a)$
   ii. $p(b)p(a|b)$
   iii. $p(b|a)p(c|b)p(a|c)$
   iv. $p(\mu)p(\sigma)\prod_{i=1}^{n} p(y_i|\mu, \sigma)$
   v. $p(\theta)p(\phi)p(y|\theta)$

5. Consider the following model for failure time data. The are n similar but not identical pieces of equipments (pumps) in a factory. For $i = 1, 2, \ldots, n$, pump i is run for a total time t_i, and incurs y_i failures. We suppose that $y_i \sim \text{Poisson}(\theta_i t_i)$. We put a prior Gamma$(\alpha, \beta)$ on the θ_i. Finally, β is modelled as Gamma(γ, δ). We treat α, γ and δ (and the $\{t_i\}$ of course) as known constants. Write down all necessary (conditional) independence assumptions you would make, not stated above, and hence write down the joint distribution of all random variables $(\beta, \{\theta_i\}, \{y_i\})$. Draw the corresponding DAG.

6. Suppose that conditional on θ, x and y are independent Bernoulli(θ), and that θ is random. Find $P\{x = 1\}$, $P\{y = 1\}$ and $P\{x = 1, y = 1\}$ in terms of $E(\theta)$ and $E(\theta^2)$. Hence show that $P\{x = 1, y = 1\} \geq P\{x = 1\} \times P\{y = 1\}$. Now consider two 0/1 random variables (w, z) such that $P\{w = 0, z = 0\} = P\{w = 1, z = 1\} = 0.1$ and $P\{w = 0, z = 1\} = P\{w = 1, z = 0\} = 0.4$ Can w and z be represented as conditionally independent given some other variable? Contrast this with de Finetti's theorem: do you see why we have to say *infinitely* exchangeable in section 4.5?