HW3, Bayesian Modelling B 2016/17

Jonathan Rougier
School of Mathematics
University of Bristol UK

In the homeworks, questions with marks are officially 'exam-style', although you can expect any homework question to appear as an exam question, unless it is explicitly 'not examinable'.

Hand in Q2 and Q4.

1. (a) Consider the prior distribution $\sigma^2 \sim \text{Gamma}(0.001, 0.001)$, where the two parameters are the shape and rate parameters, respectively. Show that under this distribution, $\log \sigma^2$ has an approximately uniform distribution on $\mathbb{R}$. [10 marks] (In an exam you would be given the PDF of a Gamma distribution.)

   Answer. The Gamma distribution for σ^2 has kernel

   $$f_{\sigma^2}(s^2) \propto (s^2)^{\alpha-1} \cdot e^{-\beta s^2}, \quad s^2 > 0,$$

   where α is the shape parameter and β is the rate parameter, so $(\alpha, \beta) = (0.001, 0.001)$ gives a kernel which is approximately $(s^2)^{-1}$ on $(0, \infty)$.

   Let $U = \log \sigma^2$, or $u = \log s^2$ in terms of arguments. According to the Transformation formula,

   $$f_U(u) = f_{\sigma^2}(s^2) \cdot \left| \frac{\mathrm{d}s^2}{\mathrm{d}u} \right| \propto \frac{1}{e^u} \cdot e^u = 1.$$

   Hence $\log \sigma^2$ is approximately uniform on $\mathbb{R}$.

   (b) Show that the following two statements are equivalent:

   i. $\mathrm{p}(x, y) \propto \mathbb{1}(x \in \mathcal{X} \wedge y \in \mathcal{Y})$.
   ii. $X \perp\!\!\!\perp Y$ and X and Y are marginally uniformly distributed.

   [5 marks]

   Answer.

   (i) implies (ii). $\mathbb{1}(x \in \mathcal{X} \wedge y \in \mathcal{Y}) = \mathbb{1}(x \in \mathcal{X}) \cdot \mathbb{1}(y \in \mathcal{Y})$, and hence $p(x, y) \propto g(x) \cdot h(y)$, which is equivalent to $X \perp\!\!\!\perp Y$. Marginalizing out X shows that $p(y) \propto \mathbb{1}(y \in \mathcal{Y})$, hence Y is marginally uniform. The same argument works for X.

   (ii) implies (i). Same argument in reverse:

   $$\begin{aligned} p(x, y) &\propto \mathbb{1}(x \in \mathcal{X}) \cdot \mathbb{1}(y \in \mathcal{Y}) \quad \text{by hypothesis} \\ &= \mathbb{1}(x \in \mathcal{X} \wedge y \in \mathcal{Y}). \end{aligned}$$

Pause for thought: (i) is a ‘rectangular’ distribution. It is what we often specify at the bottom of a hierarchical model, for example

$$\begin{aligned} \mu &\sim \mathrm{N}(0, 1000^2) \\ \sigma^2 &\sim \mathrm{Ga}(0.001, 0.001), \end{aligned}$$

which is approximately rectangular in $(\mu, \log \sigma^2)$.

2. Consider the following joint distribution for $X = (X_1, \dots, X_5)$:

$$p(x) = p(x_1) \cdot \mathrm{p}(x_2 \mid x_1) \cdot \mathrm{p}(x_3) \cdot \mathrm{p}(x_4 \mid x_1, x_3) \cdot \mathrm{p}(x_5 \mid x_3).$$

(a) Draw the DAG of f_X.

Answer.

$$\begin{tikzcd}
 & X_5 & \\
X_2 & X_3 \arrow[u] \arrow[r] & X_4 \\
X_1 \arrow[u] \arrow[urr] & &
\end{tikzcd}$$

When drawing a DAG, try to be consistent in ordering the vertices on the page. You can see that I tend to go up-and-to-the-right in my vertices. It makes it a lot easier to see what is going on. Remember that a DAG is always drawn with respect to a specific ordering of the random quantities.

(b) Draw the CIG of f_X.

Answer.

$$\begin{tikzcd}
 & X_5 & \\
X_2 & X_3 \arrow[u, dash] \arrow[r, dash] & X_4 \\
X_1 \arrow[u, dash] \arrow[ur, dash] \arrow[urr, dash] & &
\end{tikzcd}$$

which has gained an edge between X_1 and X_3 because both X_1 and X_3 are ‘parents’ of X_4.

(c) Answer True or False to the following statements:

i. $X_2 \perp\!\!\!\perp X_3, X_4 \mid X_1$

ii. $X_1 \perp\!\!\!\perp X_2 \mid X_3, X_4$

iii. $X_5 \perp\!\!\!\perp X_1, X_4 \mid X_3$

iv. $X_4 \perp\!\!\!\perp X_1, X_2 \mid X_3, X_5$

For each statement, state whether or not you could provide an answer directly from the DAG (i.e. without constructing the CIG).

Answer.

i. True. Not inferrable from the DAG, because $2 < \{3, 4\}$.

ii. False. Not inferrable from the DAG, because $1 < \{2, 3, 4\}$.

iii. True. Inferrable from the DAG. We have $X_5 \perp\!\!\!\perp X_1, X_2, X_4 \mid X_3$, and this implies $X_5 \perp\!\!\!\perp X_1, X_4 \mid X_3$.

iv. False. Not inferrable from the DAG, because $4 < 5$.

(d) Draw the DAG of (X_2, X_3, X_4, X_5), i.e. after marginalizing over X_1.

Answer. Here is the long answer, using algebra. X_1 is going to be marginalized over, so we need the factorization of $\mathrm{p}(x_2, x_3, x_4, x_5)$. For X_2 we just have $\mathrm{p}(x_2)$. For X_3:

$$\begin{aligned}
\mathrm{p}(x_3 \mid x_2) &= \frac{\mathrm{p}(x_2, x_3)}{\mathrm{p}(x_2)} \\
&= \frac{\sum_{x_1} \mathrm{p}(x_1, x_2, x_3)}{\sum_{x_1} \mathrm{p}(x_1, x_2)} \\
&= \frac{\sum_{x_1} \mathrm{p}(x_1)\, \mathrm{p}(x_2 \mid x_1)\, \mathrm{p}(x_3)}{\sum_{x_1} \mathrm{p}(x_1)\, \mathrm{p}(x_2 \mid x_1)} \\
&= \frac{\mathrm{p}(x_3) \sum_{x_1} \mathrm{p}(x_1)\, \mathrm{p}(x_2 \mid x_1)}{\sum_{x_1} \mathrm{p}(x_1)\, \mathrm{p}(x_2 \mid x_1)} \\
&= \mathrm{p}(x_3).
\end{aligned}$$

which depends on x_3 but not on x_2. For X_4,

$$\begin{aligned}
\mathrm{p}(x_4 \mid x_2, x_3) &= \frac{\mathrm{p}(x_2, x_3, x_4)}{\mathrm{p}(x_2, x_3)} \\
&= \frac{\sum_{x_1} \mathrm{p}(x_1, x_2, x_3, x_4)}{\sum_{x_1} \mathrm{p}(x_1, x_2, x_3)} \\
&= \frac{\sum_{x_1} \mathrm{p}(x_1)\, \mathrm{p}(x_2 \mid x_1)\, \mathrm{p}(x_3)\, \mathrm{p}(x_4 \mid x_1, x_3)}{\sum_{x_1} \mathrm{p}(x_1)\, \mathrm{p}(x_2 \mid x_1)\, \mathrm{p}(x_3)} \\
&= \frac{\sum_{x_1} \mathrm{p}(x_1)\, \mathrm{p}(x_2 \mid x_1)\, \mathrm{p}(x_4 \mid x_1, x_3)}{\sum_{x_1} \mathrm{p}(x_1)\, \mathrm{p}(x_2 \mid x_1)}
\end{aligned}$$

which depends on (x_2, x_3). For X_5,

$$\begin{aligned}
\mathrm{p}(x_5 \mid x_2, x_3, x_4) &= \frac{\mathrm{p}(x_2, x_3, x_4, x_5)}{\mathrm{p}(x_2, x_3, x_4)} \\
&= \frac{\sum_{x_1} \mathrm{p}(x_1, x_2, x_3, x_4, x_5)}{\sum_{x_1} \mathrm{p}(x_1, x_2, x_3, x_4)} \\
&= \frac{\sum_{x_1} \mathrm{p}(x_1)\, \mathrm{p}(x_2 \mid x_1)\, \mathrm{p}(x_3)\, \mathrm{p}(x_4 \mid x_1, x_3)\, \mathrm{p}(x_5 \mid x_3)}{\sum_{x_1} \mathrm{p}(x_1)\, \mathrm{p}(x_2 \mid x_1)\, \mathrm{p}(x_3)\, \mathrm{p}(x_4 \mid x_1, x_3)} \\
&= \frac{\mathrm{p}(x_5 \mid x_3) \sum_{x_1} \mathrm{p}(x_1)\, \mathrm{p}(x_2 \mid x_1)\, \mathrm{p}(x_4 \mid x_1, x_3)}{\sum_{x_1} \mathrm{p}(x_1)\, \mathrm{p}(x_2 \mid x_1)\, \mathrm{p}(x_4 \mid x_1, x_3)} \\
&= \mathrm{p}(x_5 \mid x_3)
\end{aligned}$$

which depends on x_3 but not on (x_2, x_4). Hence

$$\mathrm{p}(x_2, x_3, x_4, x_5) = \mathrm{p}(x_2)\, \mathrm{p}(x_3)\, \mathrm{p}(x_4 \mid x_2, x_3)\, \mathrm{p}(x_5 \mid x_3),$$

or

$$
\begin{tikzcd}
 & X_5 & \\
X_2 \arrow[rr, bend right] & X_3 \arrow[u] \arrow[r] & X_4
\end{tikzcd}
$$

The DAG has gained an edge from X_2 to X_4 because X_2 and X_4 are both 'children' of X_1 in the full distribution.

3. (a) Let $X = (X_1, \ldots, X_m)$. Give definitions for the DAG of f_X and the CIG of f_X. State and prove the Moralization Theorem. [15 marks]

Answer.

For each i let $\mathrm{pa}_i \subset \{1, \ldots, i-1\}$ be the index set for which

$$X_i \perp\!\!\!\perp X_{\overline{\mathrm{pa}}_i} \mid X_{\mathrm{pa}_i}$$

in f_X, where $\overline{\mathrm{pa}}_i$ is the complement of pa_i in $\{1, \ldots, i-1\}$; pa_i and $\overline{\mathrm{pa}}_i$ may be empty sets. A DAG is a directed graph with vertices $\{X\}$, in which there is an edge from X_i to X_j exactly when $i \in \mathrm{pa}_j$.

For each i let $\mathrm{ne}_i \in \{1, \ldots, m\} \setminus i$ be the index set for which

$$X_i \perp\!\!\!\perp X_{\overline{\mathrm{ne}}_i} \mid X_{\mathrm{ne}_i}$$

in f_X, where $\overline{\mathrm{ne}}_i$ is the complement of ne_i in $\{1, \ldots, m\} \setminus i$; ne_i and $\overline{\mathrm{ne}}_i$ may be empty sets. A CIG is an undirected graph with vertices $\{X\}$ in which there is an edge from X_i to X_j exactly when $i \in \mathrm{ne}_j$.

The Moralization theorem states: to convert a DAG into a CIG,

i. Insert an edge between every pair of vertices which share a child.
ii. Replace all directed edges with undirected edges.

The proof is in the handout.

(b) State the Hammersley-Clifford Theorem, and explain its role in interpeting the CIG. [10 marks]

Answer. The HC theorem states (subject to a positivity condition, see below) that $X_A \perp\!\!\!\perp X_B \,|\, X_C$ if and only if every path on the CIG from X_A to X_C passes through X_B. It allows us to use the CIG to read off all possible conditional independence statements.

The positivity condition cannot be lifted. If $\operatorname{supp} X_A$ is the support of X_A in f_{X_A}, i.e. the set $\{x_A : f_{X_A}(x_A) > 0\}$, then the positivity condition is

$$\operatorname{supp} X = \prod_{i=1}^{m} \operatorname{supp} X_i,$$

or, to put it crudely, that the support of X is 'rectangular'. It rules out deterministic relationships between elements of X (why?).

4. Consider the 'old-fashioned' regression model

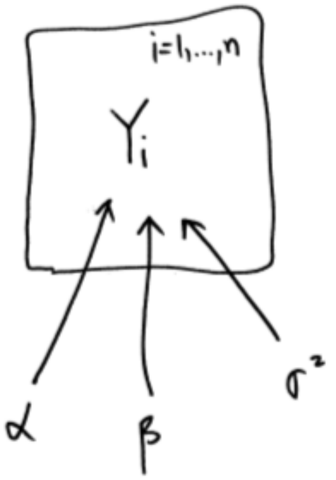
$$Y_i = \alpha + \beta\, X_i + \epsilon_i, \qquad i = 1, \ldots, n,$$

where $\epsilon \overset{\text{iid}}{\sim} \text{Normal}(0, \sigma^2)$.

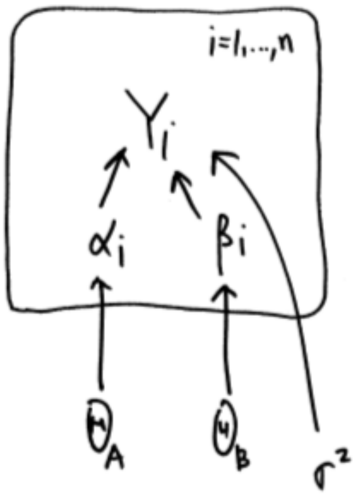
(a) Write this model as a DAG, using a plate. Hint: α, β, and σ^2 are parameters but ϵ is not. [5 marks]

(b) Generalize this DAG so that each case gets its own (α_i, β_i), where the α's and β's are each exchangeable. [5 marks]

Answer. For (a) and (b):



'old fashioned'



'exchangeable α and β'

(c) Make explicit choices for the marginal and conditional distributions in the DAG, and identify the restriction that forces the generalized model to behave like the old-fashioned one. [5 marks]

Answer. I'll write the model in extensive form:

$$\begin{aligned}
Y_i \mid \alpha_i, \beta_i, \sigma^2 &\sim \text{N}(\alpha_i + \beta_i\, X_i, \sigma^2) & i &= 1, \ldots, n \\
\alpha_i \mid \mu_A, \tau_A^2 &\sim \text{N}(\mu_A, \tau_A^2) & i &= 1, \ldots, n \\
\beta_i \mid \mu_B, \tau_B^2 &\sim \text{N}(\mu_B, \tau_B^2) & i &= 1, \ldots, n \\
\mu_A &\sim \text{N}(0, 1000^2) \\
\mu_B &\sim \text{N}(0, 1000^2) \\
\tau_A^2 &\sim \text{Ga}(0.001, 0.001) \\
\tau_B^2 &\sim \text{Ga}(0.001, 0.001)
\end{aligned}$$

where $\Theta_A = (\mu_A, \tau_A^2)$ and similarly for Θ_B. The restrictions are $\tau_A^2 \to 0$ and $\tau_B^2 \to 0$, which would force each α_i to the same unknown value μ_A, and each β_i to the same unknown value μ_B.