SC7/SM6 Bayes Methods

Second problem sheet for classes in week 5

1. (a) Suppose $h : \Omega \to \mathbb{R}$ is a function chosen so that the expectations below exist. Suppose we have two models: model $m = 1$ is $\pi_1(\theta), p_1(y|\theta), p_1(y)$; model $m = 2$ is $\pi_2(\phi)p_2(y|\phi), p_2(y)$. Suppose the parameter spaces are the same so that for example

$$
\int_{\Omega} \pi_1(\theta)d\theta = \int_{\Omega} \pi_2(\phi)d\phi = 1.
$$

Show that the Bayes factor $B_{1,2} = p_1(y)/p_2(y)$ can be written

$$
\frac{p_1(y)}{p_2(y)} = \frac{E_{\phi|y,m=2}(\pi_1(\phi)p_1(y|\phi)h(\phi))}{E_{\theta|y,m=1}(\pi_2(\theta)p_2(y|\theta)h(\theta))}
$$

[Please assume all the random variables have probability densities]

(b) Suppose now the two models have different numbers of parameters. To be concrete suppose $\theta \in \mathbb{R}^p$ and $\phi = (\theta, \psi)$ with $\psi \in \mathbb{R}^q$, so that $\phi \in \mathbb{R}^{p+q}$ and

$$
\int_{\mathbb{R}^{p+q}} \pi_2(\phi)d\phi = \int_{\mathbb{R}^p} \left[ \int_{\mathbb{R}^q} \pi_2(\theta, \psi)d\psi \right] d\theta = 1,
$$

where in our notation $\pi_2(\theta, \psi) = \pi_2(\phi)$ when $\phi = (\theta, \psi)$. We want to compare the model $\pi_1(\theta), p_1(y|\theta), p_1(y)$ with $\pi_2(\phi)p_2(y|\phi), p_2(y)$. Explain briefly why

$$
\frac{p_1(y)}{p_2(y)} \neq \frac{E_{\phi|y,m=2}(\pi_1(\phi)p_1(y|\phi)h(\phi))}{E_{\theta|y,m=1}(\pi_2(\theta)p_2(y|\theta)h(\theta))}
$$

as the RHS is nonsensical as it stands.

(c) Let $Q(\psi)$ be the prior for the $\psi$ parameters in model $m = 2$. Show that

$$
\frac{p_1(y)}{p_2(y)} = \frac{E_{(\theta, \psi)|y,m=2}(Q(\psi)\pi_1(\theta)p_1(y|\theta)h(\theta, \psi))}{E_{(\theta, \psi)|y,m=1}(\pi_2(\theta, \psi)p_2(y|\theta, \psi)h(\theta, \psi))}
$$

where $\psi \sim Q$ in the expectation in the denominator.

(d) Comment briefly on how this last identity may be used for model comparison for models defined on spaces of unequal dimension.\(^1\)

2. In the radiocarbon dating example, suppose the dated materials are found in layers piled up on one another, with $y_{i,j}$ the radiocarbon date for $\theta_{i,j}$, the $j$th date in the $i$th layer. Let $L < \psi_1 < \psi_2 < ... < \psi_M < U$ be the age parameters for the layer boundaries. If we have $n_i$ dates from the $i$th stratum we know that for $i = 1, 2, ..., M-1$, and $j = 1, 2, ..., n_i$, $\psi_i < \theta_{i,j} < \psi_{i+1}$. Let $\psi = (\psi_1, ..., \psi_M)$ and $\theta = (\theta_1, ..., \theta_{M-1})$ with $\theta_i = (\theta_{i,1}, ..., \theta_{i,n_i})$.

Justify the prior density \( \pi_3(\theta, \psi) \) with reference to the principles of prior elicitation:

\[
\pi_3(\theta, \psi) \propto \frac{1}{(\psi_M - \psi_1)^{M-2}} \left( \frac{U - L - (\psi_M - \psi_1)}{(\psi + 1 - \psi_1)} \right)^{M-1} \prod_{i=1}^{M-1} \left( \frac{1}{\psi_i + 1 - \psi_1} \right)^{n_i}.
\]

Hint: how are the layer boundary dates \( \psi_2, ..., \psi_{M-1} \) generated?

3. For \( x_i \in \{0, 1\} \) for \( i = 1, 2, 3, ... \), the sequence of distributions \( p_n(x_1, ..., x_n), n = 1, 2, 3, ... \) is *marginally consistent* if

\[
p_n(x_1, ..., x_n) = p_{n+1}(x_1, ..., x_n, 0) + p_{n+1}(x_1, ..., x_n, 1).
\]

John elicits a separate prior \( p_n(x_1, x_2, ..., x_n) \) for each value of \( n = 1, 2, ... \). John’s priors are not marginally consistent, that is

\[
p_n(x_1, ..., x_n) \neq p_{n+1}(x_1, ..., x_n, 0) + p_{n+1}(x_1, ..., x_n, 1).
\]

(a) (optional) Can you think of a well-known family of distributions on binary random variables that are not marginally consistent? *Hint, a model for binary images...*

(b) Show that John’s priors do not satisfy the Savage axioms (consider the first three).

(c) Show that, under John’s priors, \( x_1, x_2, x_3, ... \) do not form an infinite exchangeable sequence. *Hint: show infinite exchangeable sequences are marginally consistent.*

4. Suppose \( x_1, x_2, x_3, ... \) is now an infinite exchangeable sequence of binary random variables.

(a) Show that \( \text{cov}(x_i, x_j) \geq 0 \) for all \( i, j \in \{1, 2, 3, ...\} \).

(b) Construct a prior for \( x_1, x_2, x_3, ... \) representing the following prior knowledge: (A) the variables are exchangeable; (B) \( \Pr(x_i = 1) = p \); and (C) \( \text{var}(\bar{x}) = v \) where \( \bar{x} = n^{-1} \sum_{i=1}^{n} x_i \) and \( 0 \leq p \leq 1 \) and \( v \) are prior parameters we wish to specify separately. Note any constraints on \( p \) and \( v \).

*Hint: this is just Q1 of sheet 1 again.*

5. Let \( X_1, X_2 \) be binary random variables. Table entries below give probabilities, \( p(x_1, x_2) = \Pr(X_1 = x_1, X_2 = x_2) \), for outcomes \( (X_1, X_2) = (x_1, x_2) \) indicated by row and column.

<table>
<thead>
<tr>
<th></th>
<th>( X_1 = 0 )</th>
<th>( X_1 = 1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( X_2 = 0 )</td>
<td>0</td>
<td>1/2</td>
</tr>
<tr>
<td>( X_2 = 1 )</td>
<td>1/2</td>
<td>0</td>
</tr>
</tbody>
</table>

(a) Show that \( X_1 \) and \( X_2 \) are exchangeable.

(b) Show that there does not exist a distribution \( F \) such that

\[
p(x_1, x_2) = \int_0^1 \prod_{i=1, 2} x_i^{x_i} (1 - p)^{1-x_i} dF(p).
\]

\(^2\text{ie, de Finetti’s theorem need not hold if the sequence is finite. This example from P. Diaconis and Freedman D. (1980). Finite Exchangeable Sequences. Ann. Probab. v8 p745–764.}\)