Computational Statistics
9. exercise session, 23.4.2010

We will investigate several techniques in the context of the following case-control study. The table gives data on the relationship of coffee drinking and pancreatic cancer

	Cases	Controls
≥ 0 cups of coffee per day	347	555
0 cups of coffee per day	20	88
n_i	367	643

The cases were identified from hospital records and the controls were sampled from the patient populations of physicians participating in the study. The aim is to study whether the exposure (regular coffee consumption) is associated with the disease. Let E and D denote the events that a randomly selected individual from the study base is exposed and diseased, respectively. Then the columns of the table carry information on the conditional probabilities $P(E \mid D)$ and $P(E \mid D^c)$. If these are equal, then the events E and D are easily shown to be independent.

Let Y_{ij} be the count on the ith row and jth column of this 2×2 matrix. The likelihood corresponds to the assumptions

$$[Y_{11} \mid \pi_1, \pi_2] \sim \mathrm{Bin}(n_1, \pi_1), \qquad [Y_{12} \mid \pi_1, \pi_2] \sim \mathrm{Bin}(n_2, \pi_2)$$

independently. Here n_i is the sum of the ith column (n_1 and n_2 are assumed fixed quantities), and the interpretations of the parameters are as follows:

$$\pi_1 = P(E \mid D), \qquad \pi_2 = P(E \mid D^c).$$

Under the null hypothesis H_0, the exposure and disease status of a person are independent, and therefore $\pi_1 = \pi_2 = \pi$ and we take the prior of π to be $\mathrm{Be}(\alpha_0, \beta_0)$. Under the alternative hypothesis H_1 we have two separate parameters π_1 and π_2 which we take to have independent $\mathrm{Be}(\alpha_0, \beta_0)$ distributions in the prior. We choose $\alpha_0 = \beta_0 = 1$.

In order to make the situation compatible with the lectures, we label hypothesis H_0 as model 0 and hypothesis H_1 as model 1.

1. In a certain two-model situation (model 0 versus model 1) the Bayes factor in favor of model 1 turns out to be 376.6.

a) Calculate the posterior model probabilities, when we assume that the prior model probabilities are equal.

b) How would we have to set the prior model probabilities, in order to obtain equal posterior model probabilities?

2. Calculate marginal likelihoods for the two hypotheses in the case-control study, and calculate the Bayes factor in favor of the alternative hypothesis. The likelihood and the prior are conjugated, both under H_0 and under H_1, and therefore all these quantities can be calculated analytically. (Of course, such analytical solution would be impossible in a more complicated situation.)

3. Calculate the marginal likelihoods and the Bayes factor for the case-control study under H_0 and under H_1 using Laplace approximation, formula (10.7).

4. Calculate BIC values for H_0 and H_1 in the case-control study (using $n = n_1 + n_2$), and estimate the posterior model probabilities using the approximation (10.12).

5. Consider RJMCMC in the situation, where

$$\begin{aligned} \theta_1 &= \pi \in (0,1), && \text{(model 1)} \\ \theta_2 &= (\pi_1, \pi_2) \in (0,1) \times (0,1), && \text{(model 2).} \end{aligned}$$

The move $1 \to 2$ is effected by the following steps,

- convert π to log-odds scale: $\psi \leftarrow \log(\pi/(1-\pi))$.
- draw u from the density g;
- calculate $\psi_1 \leftarrow \psi - u$ and $\psi_2 \leftarrow \psi + u$;
- convert the two proposed log-odds quantities to probability scale:

$$\pi_1 \leftarrow \frac{\mathrm{e}^{\psi_1}}{1+\mathrm{e}^{\psi_1}}, \qquad \pi_2 \leftarrow \frac{\mathrm{e}^{\psi_2}}{1+\mathrm{e}^{\psi_2}}.$$

a) What is the move $2 \to 1$?

b) Write formulas for the test ratio r for moves $1 \to 2$ and for $2 \to 1$.