Introduction to Bayesian Data Analysis

Tutorial 2

(1) Suppose that in each individual of a large population there is a pair of genes, each of which can be either x or X, that controls eye colour: those with xx have blue eyes, while heterozygotes (those with Xx or xX) and those with XX have brown eyes. The proportion of blue-eyed individuals is p^2 and of heterozygotes is $2p(1-p)$, where $0 < p < 1$. Each parent transmits one of its own genes to the child; if a parent is a heterozygote, the probability that it transmits the gene of type X is $\frac{1}{2}$. Assuming random mating, show that among brown-eyed children of brown-eyed parents, the expected proportion of heterozygotes is $2p/(1+2p)$. Suppose Judy, a brown-eyed child of brown-eyed parents, marries a heterozygote, and they have n children, all brown-eyed. Find the posterior probability that Judy is a heterozygote and the probability that her first grandchild has blue eyes.

(2) Problem 2.3 (Hoff).
Full conditionals: Let X, Y,Z be random variables with joint density (discrete or continuous) $p(x, y, z) \propto f(x, z)g(y, z)h(z)$. Show that

(a) $p(x|y, z) \propto f(x, z)$. That is $p(x|y, z)$ is a function of x and z

(b) $p(y|x, z) \propto g(y, z)$. That is $p(y|x, z)$ is a function of y and z

(c) X and Y are conditionally independent given Z.

(3) Problem 2.5 (Hoff).
Urns: Suppose urn H is filled with 40% green balls and 60% red balls, and urn T is filled with 60% green balls and 40% red balls. Someone will flip a coin and then select a ball from urn H or urn T depending on whether the coin lands heads or tails, respectively. Let X be 1 or 0 if the coin lands heads or tails, and let Y be 1 or 0 if the ball is green or red.

   (a) Write out the joint distribution of X and Y in a table.
   (b) Find E[Y]. What is the probability that the ball is green?
   (c) Find Var[Y |X = 0], Var[Y |X = 1] and Var[Y]. Thinking of variance as measuring uncertainty, explain intuitively why one of these variances is larger than the others, in expectation.
   (d) Suppose you see that the ball is green. What is the probability that the coin turned up tails?

(4) In a forest area of Northern Europe there may be wild cats. At a particular time the number X of cats can be between 0 and 5 with

$$Pr(X = x) = \binom{5}{x} 0.6^x 0.4^{5-x} \ (x = 0, .., 5)$$

A survey is made but the wild cat is difficult to spot and, given that the number present is $X = x$, the number Y observed has a probability distribution with

$$Pr(Y = y|X = x) = \begin{cases} \binom{x}{y} 0.3^y 0.7^{x-y} & 0(\leq y \leq x) \\ 0 & x < y \end{cases}$$

Given that two wild cats are observed, find the conditional probability distribution of X. That is find, Pr(X=0 | Y=2),..., Pr(X=5 | Y=2)

(5) In a certain small town there are n taxis which are clearly numbered 1,2,...,n. Before we visit the town we do not know the value of n but our probabilities for the possible values of n are as follows:

n	0	1	2	3	4	5	6	7	8	≥ 9
prob	0.00	0.11	0.12	0.13	0.14	0.14	0.13	0.12	0.11	0.00

On a visit to the town we take a taxi which we assume would be equally likely to be any of the taxis 1,2,...,n. It is taxi number 5. Find our new probabilities for the value of n.

(6) Let $F(x, y) = x - x\log(x/y)$ for $0 < x \leq y < 1$. For (a) and (b), assume that $F(x, y)$ is a two-dimensional CDF.

(a) Find the density function $f(x, y)$, corresponding to $F(x, y)$.

(b) Find the marginal distributions of X and Y for $(X, Y) \sim F(x, y)$

(7) **Biostat Theorem**: In this problem you will prove the following result, useful in biostatistics (in survival times).

Theorem 1: Let $Y_1 = X_1/\lambda_1$ and $Y_2 = X_2/\lambda_2$ be independent (scaled) Exponentials, with $X_1, X_2 \sim Exp(1)$ and $\lambda_1, \lambda_2 > 0$ constants. Define

$$W \equiv min(Y_1, Y_2) \text{ and } B_0 \equiv I_{Y_1 < Y_2}$$

where I_A is the indicator random variable for an event A. Then $W \perp B_0$

The proof will be done in several parts. Let $U_0 \equiv X_2/(X_1 + X_2)$ and $T \equiv X_1 + X_2$.

(a) Show that $W \sim (\lambda_1 + \lambda_2)^{-1}\text{Exp}(1)$ and $B_0 \sim \text{Bern}_{\frac{\lambda_1}{\lambda_1+\lambda_2}}$ (Hint: for the latter, use the fact that $U_0 \sim Beta(1, 1)$, rather than an integral).

(b) We can prove the biostat theorem assuming the following lemma.

Lemma: Let $0 < p < 1$ be a constant and $Y \sim Unif(0, 1)$. Define

$$B \equiv I_{U \leq p} \text{ and } M \equiv \left(\frac{U}{p}, \frac{1-U}{1-p}\right)$$

Then the indicator random variable $B \sim Bern(p)$ is independent of $M \sim Unif(0, 1)$.

Prove the Lemma. Hint: Compute $P(U \leq p | M \geq m)$

(8) Exchangeable prior distributions: suppose it is known *a priori* that the $2J$ parameters $\theta_1, ..., \theta_{2J}$ are clustered into two groups, with exactly half being drawn from a N(1,1) distribution, and the other half being drawn from a N(-1,1) distribution, but we have not observed which parameters come from which distribution.

(a) Are $\theta_1, ..., \theta_{2J}$ exchangeable under this prior distribution?

(b) Show that this distribution cannot be written as a mixture of independent and identically distributed components.

(c) Why can we not simply take the limit $J \to \infty$ and get a counterexample to deFinetti's theorem?

(9) Exchangeability with known model parameters: For each of the following three examples, answer: (i) Are observations y_1 and y_2 exchangeable? (ii) Are observations y_1 and y_2 independent? (iii) Can we act as if the two observations are independent?

(a) A box has one black ball and one white ball. We pick a ball y_1 at random, put it back and pick another ball y_2 at random.

(b) A box has one black and one white ball. We pick a ball y_1 at random, we do not put it back, then we pick ball y_2.

(c) A box has a million black balls and a million white balls. We pick a ball y_1 at random, we do not put it back, then we pick y_2 at random.

(a). (i) exchangable, (ii) independent

(finite sampling with replacement)

(iii). yes, b/c they are in fact independent

(b). (i) finite sampling without replacement

exchangable,

(ii). not independent (perfectly correlated)

(iii). cannot act like they're independent

(c). (i). exchangable.

(ii). not independent

(iii). yes b/c sample size is large.

(sample from infinite population)

can treat ~~it~~ them as independent events.