Introduction to Bayes Linear Statistics: Solutions Sheet.

Jonathan Cumming and Ian Vernon

September 3, 2010

1. The following example comes from a healthcare application related to glucose tolerance tests. You will need to work to about 4dp to achieve reasonable accuracy.

   We have two vectors of random quantities $B = [B_1, B_2]^T$ and $D = [D_1, D_2]^T$ and we will adjust beliefs over B by observation of D. We therefore specify prior means, variances and covariances for each quantity as follows:

$$\mathrm{E}(B) = \begin{pmatrix} \mathrm{E}(B_1) \\ \mathrm{E}(B_2) \end{pmatrix} = \begin{pmatrix} 4 \\ 6 \end{pmatrix} \quad (1)$$

$$\mathrm{E}(D) = \begin{pmatrix} \mathrm{E}(D_1) \\ \mathrm{E}(D_2) \end{pmatrix} = \begin{pmatrix} 4 \\ 6 \end{pmatrix} \quad (2)$$

$$\mathrm{Var}(D) = \begin{pmatrix} \mathrm{Var}(D_1) & \mathrm{Cov}(D_1, D_2) \\ \mathrm{Cov}(D_2, D_1) & \mathrm{Var}(D_2) \end{pmatrix} = \begin{pmatrix} 1 & 0.7 \\ 0.7 & 2 \end{pmatrix} \quad (3)$$

$$\mathrm{Var}(B) = \begin{pmatrix} \mathrm{Var}(B_1) & \mathrm{Cov}(B_1, B_2) \\ \mathrm{Cov}(B_2, B_1) & \mathrm{Var}(B_2) \end{pmatrix} = \begin{pmatrix} 1 & 0.7 \\ 0.7 & 2 \end{pmatrix} \quad (4)$$

$$\mathrm{Cov}(B, D) = \begin{pmatrix} \mathrm{Cov}(B_1, D_1) & \mathrm{Cov}(B_1, D_2) \\ \mathrm{Cov}(B_2, D_1) & \mathrm{Cov}(B_2, D_2) \end{pmatrix} = \begin{pmatrix} 0.6 & 0.3 \\ 0.3 & 0.4 \end{pmatrix} \quad (5)$$

   (a) Univariate Case: find the adjusted expectation $\mathrm{E}_{D_1}(B_1)$ and the adjusted variance $\mathrm{Var}_{D_1}(B_1)$ for the adjustment of B_1 by D_1 only.

   (b) Multivariate Case: find the adjusted expectation vector $\mathrm{E}_D(B)$ and the adjusted variance $\mathrm{Var}_D(B)$ for the adjustment of B by D.

   (c) Find the adjusted variance $\mathrm{Var}_D(A)$ for the quantity $A = B_2 - B_1$ (i) directly, by using prior information on A itself, and (ii) indirectly, using the adjusted variances and covariances for B_1 and B_2 found in part (b). Verify that in this case you get the same results.

   Solution 1.(a) Using the formula for the adjusted expectation from the notes, we have:

$$\mathrm{E}_{D_1}(B_1) = \mathrm{E}(B_1) + \mathrm{Cov}(B_1, D_1)\mathrm{Var}(D_1)^{-1}(D_1 - \mathrm{E}(D_1)) \quad (6)$$

$$
\begin{aligned}
&= 4 + 0.6 \times 1^{-1} \times (D_1 - 4) && (7)\\
&= 0.6D_1 + 1.6 && (8)
\end{aligned}
$$

and similarly for the adjusted variance:

$$
\begin{aligned}
\mathrm{Var}_{D_1}(B_1) &= \mathrm{Var}(B_1) - \mathrm{Cov}(B_1, D_1)\mathrm{Var}(D_1)^{-1}\mathrm{Cov}(D_1, B_1) && (9)\\
&= 1 - 0.6 \times 1^{-1} \times 0.6 && (10)\\
&= 0.64 && (11)
\end{aligned}
$$

Solution 1.(b) Again using the formula for the adjusted expectation from the notes, we have:

$$
\begin{aligned}
\mathrm{E}_D(B) &= \mathrm{E}(B) + \mathrm{Cov}(B, D)\mathrm{Var}(D)^{-1}(D - \mathrm{E}(D)) && (12)\\
&= \begin{pmatrix} 4 \\ 6 \end{pmatrix} + \begin{pmatrix} 0.6 & 0.3 \\ 0.3 & 0.4 \end{pmatrix} \begin{pmatrix} 1 & 0.7 \\ 0.7 & 2 \end{pmatrix}^{-1} \left(\begin{pmatrix} D_1 \\ D_2 \end{pmatrix} - \begin{pmatrix} 4 \\ 6 \end{pmatrix} \right) && (13)\\
&= \begin{pmatrix} 0.656D_1 - 0.079D_2 + 1.85 \\ 0.212D_1 - 0.126D_2 + 4.40 \end{pmatrix} && (14)
\end{aligned}
$$

and similarly for the adjusted variance:

$$
\begin{aligned}
\mathrm{Var}_D(B) &= \mathrm{Var}(B) - \mathrm{Cov}(B, D)\mathrm{Var}(D)^{-1}\mathrm{Cov}(D, B) && (15)\\
&= \begin{pmatrix} 1 & 0.7 \\ 0.7 & 2 \end{pmatrix} - \begin{pmatrix} 0.6 & 0.3 \\ 0.3 & 0.4 \end{pmatrix} \begin{pmatrix} 1 & 0.7 \\ 0.7 & 2 \end{pmatrix}^{-1} \begin{pmatrix} 0.6 & 0.3 \\ 0.3 & 0.4 \end{pmatrix} && (16)\\
&= \begin{pmatrix} 0.631 & 0.535 \\ 0.525 & 1.89 \end{pmatrix} && (17)
\end{aligned}
$$

Solution 1.(c) part (i) We obtain the prior expectation and variance of A as follows:

$$
\begin{aligned}
\mathrm{E}(A) &= \mathrm{E}(B_2) - \mathrm{E}(B_1) = 6 - 4 = 2 && (18)\\
\mathrm{Var}(A) &= \mathrm{Var}(B_2 - B_1) && (19)\\
&= \mathrm{Var}(B_2) + \mathrm{Var}(B_1) - 2\mathrm{Cov}(B_2, B_1) && (20)\\
&= 2 + 1 - 2 \times 0.7 && (21)\\
&= 1.6 && (22)
\end{aligned}
$$

and the prior covariance of A with D is given by:

$$
\begin{aligned}
\mathrm{Cov}(A, D) &= \mathrm{Cov}(B_2 - B_1, D) && (23)\\
&= \mathrm{Cov}(B_2, D) - \mathrm{Cov}(B_1, D) && (24)\\
&= (0.3, 0.4) - (0.6, 0.3) = (-0.3, 0.1) && (25)
\end{aligned}
$$

We can now update the variance of A as usual:

$$\begin{aligned}
\mathrm{Var}_D(A) &= \mathrm{Var}(A) - \mathrm{Cov}(A, D)\mathrm{Var}(D)^{-1}\mathrm{Cov}(D, A) && (26)\\
&= 1.6 - (-0.3, 0.1)\begin{pmatrix} 1 & 0.7 \\ 0.7 & 2 \end{pmatrix}^{-1}\begin{pmatrix} -0.3 \\ 0.1 \end{pmatrix} && (27)\\
&= 1.45 && (28)
\end{aligned}$$

Solution 1.(c) part (ii) We can instead use the updated variance and covariances of B_1 and B_2:

$$\begin{aligned}
\mathrm{Var}_D(A) &= \mathrm{Var}_D(B_2 - B_1) && (29)\\
&= \mathrm{Var}_D(B_2) + \mathrm{Var}_D(B_1) - 2\mathrm{Cov}_D(B_2, B_1) && (30)\\
&= 1.89 + 0.631 - 2 \times 0.535 && (31)\\
&= 1.45 && (32)
\end{aligned}$$

The same as for part (i). The second line of this calculation just comes from the linearity of the adjusted variance.

2. Prove the following:

   (a) For any r-vectors B, C, constants a_1 and a_2, and vector D,

   $$\mathrm{E}_D(a_1 B + a_2 C) = a_1 \mathrm{E}_D(B) + a_2 \mathrm{E}_D(C). \quad (33)$$

   (b) For any vectors B, D,

   $$\mathrm{E}(\mathrm{E}_D(B)) = \mathrm{E}(B). \quad (34)$$

   (c) $\mathrm{Cov}(\mathrm{E}_D(B), B - \mathrm{E}_D(B)) = 0$.

   (d) The adjusted expectation $\mathrm{E}_D(X)$ is given by the linear combination $\sum_i h_i D_i$, where the h_i are chosen to minimise the following quadratic penalty score:

   $$L = c[X - \sum_{i=1}^{k} h_i D_i]^2 \quad (35)$$

   If $D = (D_1, ..., D_k)$ is the vector of indicator functions for a partition, i.e. one and only one of the quantities D_i will take value 1, and the rest will take the value 0, show that L can be rewritten as:

   $$L = \sum_{i=1}^{k} c D_i (X - h_i)^2 \quad (36)$$

which is the form of a"called off bet", and defines the conditional expectation $\mathrm{E}(X|D)$. Hence show that, if D is a partition:

$$\mathrm{E}_D(X) = \sum_{i=1}^{k} \mathrm{E}(X|D_i)D_i. \tag{37}$$

and the adjusted expectation equals the conditional expectation.

Solution 2.(a) To prove linearity we use the adjusted expectation formula:

$$\begin{aligned}
& \mathrm{E}_D(a_1B + a_2C) = \\
= \;& \mathrm{E}(a_1B + a_2C) + \mathrm{Cov}(a_1B + a_2C, D)\mathrm{Var}(D)^{-1}(D - \mathrm{E}(D)) \\
= \;& a_1\mathrm{E}(B) + a_2\mathrm{E}(C) + (a_1\mathrm{Cov}(B, D) + a_2\mathrm{Cov}(C, D))\mathrm{Var}(D)^{-1}(D - \mathrm{E}(D)) \\
= \;& a_1(\mathrm{E}(B) + \mathrm{Cov}(B, D)\mathrm{Var}(D)^{-1}(D - \mathrm{E}(D))) + \\
& a_2(\mathrm{E}(C) + \mathrm{Cov}(C, D)\mathrm{Var}(D)^{-1}(D - \mathrm{E}(D))) \\
= \;& a_1\mathrm{E}_D(B) + a_2\mathrm{E}_D(C) \qquad \text{as required.}
\end{aligned}$$

Solution 2.(b) Trivially we have:

$$\begin{aligned}
\mathrm{E}(\mathrm{E}_D(B)) &= \mathrm{E}(\mathrm{E}(B) + \mathrm{Cov}(B, D)\mathrm{Var}(D)^{-1}(D - \mathrm{E}(D))) && (38) \\
&= \mathrm{E}(B) + \mathrm{Cov}(B, D)\mathrm{Var}(D)^{-1}(\mathrm{E}(D) - \mathrm{E}(D)) && (39) \\
&= \mathrm{E}(B) \qquad \text{as required.} && (40)
\end{aligned}$$

Solution 2.(c) We have that:

$$\begin{aligned}
& \mathrm{Cov}[\ \mathrm{E}_D(B)\ ,\ B - \mathrm{E}_D(B)\ ] = && (41) \\
= \;& \mathrm{Cov}[\ \mathrm{E}(B) + \mathrm{Cov}(B, D)\mathrm{Var}(D)^{-1}(D - \mathrm{E}(D))\ , && (42) \\
& \qquad B - \mathrm{E}(B) - \mathrm{Cov}(B, D)\mathrm{Var}(D)^{-1}(D - \mathrm{E}(D))\ ] && (43) \\
= \;& \mathrm{Cov}[\ \mathrm{Cov}(B, D)\mathrm{Var}(D)^{-1}D\ ,\ B - \mathrm{Cov}(B, D)\mathrm{Var}(D)^{-1}D\ ] && (44) \\
= \;& \mathrm{Cov}(B, D)\mathrm{Var}(D)^{-1}\ \ \mathrm{Cov}[\ D\ ,\ B - \mathrm{Cov}(B, D)\mathrm{Var}(D)^{-1}D\ ] && (45) \\
= \;& \mathrm{Cov}(B, D)\mathrm{Var}(D)^{-1}(\mathrm{Cov}[D, B] - \mathrm{Cov}[D, D](\mathrm{Cov}(B, D)\mathrm{Var}(D)^{-1})^T && (46) \\
= \;& \mathrm{Cov}(B, D)\mathrm{Var}(D)^{-1}(\mathrm{Cov}(D, B) - \mathrm{Var}(D)\mathrm{Var}(D)^{-1}\mathrm{Cov}(D, B)) && (47) \\
= \;& 0 \qquad \text{as required.} && (48)
\end{aligned}$$

Solution 2.(d) As D is a partition, we have that $D_i \in \{0, 1\}$ and therefore $D_i^2 = D_i$, and also $\sum_i D_i = 1$. We can expand L as:

$$L \;=\; c[X - \sum_{i=1}^{k} h_iD_i]^2 \tag{49}$$

$$= c[X^2 - 2X\sum_{i=1}^{k} h_iD_i + (\sum_{i=1}^{k} h_iD_i)^2] \quad (50)$$

$$= c[X^2\sum_{i=1}^{k} D_i - 2X\sum_{i=1}^{k} h_iD_i + \sum_{i=1}^{k} h_i^2D_i] \quad (51)$$

$$= \sum_{i=1}^{k} cD_i(X^2 - 2Xh_i + h_i^2) \quad (52)$$

$$= \sum_{i=1}^{k} cD_i(X - h_i)^2 \quad (53)$$

as required. Hence to minimise L we choose h_i that are actually the conditional expectations $h_i = \mathrm{E}(X|D_i)$ and therefore the adjusted expectation is exactly equal to the conditional expectation as:

$$\mathrm{E}_D(X) = \sum_{i=1}^{k} h_iD_i = \sum_{i=1}^{k} \mathrm{E}(X|D_i)D_i \quad (54)$$

3. For the example in question 1, suppose that we observe the values $d_1 = 5.4$ and $d_2 = 9.8$.

   (a) Using the standardised change $S(d_i)$, assess whether the data d_1 and d_2 are discrepant.

   (b) Find the observed adjusted expectation $\mathrm{E}_{d_1}(B_1)$ of B_1 adjusted by d_1 only (you might want to use the results of Q1(a)).

   (c) Find the observed adjusted expectation vector $\mathrm{E}_d(B)$ of B adjusted by d. Why is it unnecessary to calculate the adjusted variance?

Solution 3.(a) Using the formula for $S(d_i)$ from the notes we have:

$$S(d_1) = \frac{d_1 - \mathrm{E}(D_1)}{\sqrt{\mathrm{Var}(D_1)}} = \frac{5.4-4}{\sqrt{1}} = 1.4 \quad (55)$$

$$S(d_2) = \frac{d_2 - \mathrm{E}(D_2)}{\sqrt{\mathrm{Var}(D_2)}} = \frac{9.8-6}{\sqrt{2}} = 2.69 \quad (56)$$

Hence we would consider d_2 to be borderline acceptable, but that it possibly warrants further investigation.

Solution 3.(b) Using the formula for $\mathrm{E}_{D_1}(B_1)$ obtained in question 1.(a) and replacing D_1 with $d_1 = 5.4$ gives:

$$\mathrm{E}_{d_1}(B_1) = 4.84 \quad (57)$$

Solution 3.(c) Using the formula for $\mathrm{E}_D(B)$ obtained in question 1.(b) and replacing D with $d = (5.4, 9.8)^T$ gives:

$$\mathrm{E}_d(B) = \begin{pmatrix} 4.62 \\ 6.77 \end{pmatrix} \tag{58}$$

4. The Bearing $\mathbb{Z}_d(B)$ is the linear combination of B_1 and B_2 that possesses the largest squared change in expectation, relative to prior variance and is given by:

$$\mathbb{Z}_d(B) = [\mathrm{E}_d(B) - \mathrm{E}(B)]^T \mathrm{Var}(B)^{-1}[B - \mathrm{E}(B)] \tag{59}$$

The size of this maximum squared change in expectation, along the direction of the bearing, is denoted $\mathrm{Size}_d(B)$ and given by:

$$\mathrm{Size}_d(B) = [\mathrm{E}_d(B) - \mathrm{E}(B)]^T \mathrm{Var}(B)^{-1}[\mathrm{E}_d(B) - \mathrm{E}(B)] \tag{60}$$

(a) For the example in question 1, find and interpret the bearing $\mathbb{Z}_d(B)$ for the adjustment.

(b) Calculate the size of the adjustment $\mathrm{Size}_d(B)$.

Solution 4.(a) Using $\mathrm{E}_d(B)$ as found in question 3.(c) and inserting it into equation (59) along with the priors for B gives the bearing as:

$$\mathbb{Z}_d(B) = 0.457B_1 + 0.228B_2 + 3.19 \tag{61}$$

This implies that we adjust our expectation of B_1 substantially more than that of B_2.

Solution 4.(b) Again using $\mathrm{E}_d(B)$ as found in question 3.(c) and now inserting it into equation (60) along with the priors for B gives the size to be:

$$\mathrm{Size}_d(B) = 0.458 \tag{62}$$

This implies that the squared change in expectation in any direction is less than or equal to 0.458, and that along the direction of the bearing, we adjust our expectation by $0.677 = \sqrt{0.458}$ prior standard deviations. We hence adjust our beliefs by a moderate amount.