Elementary hypothesis testing

← Back to Chapters

Exercise 4.1

Given

$$P(D_1...D_m|H_iX) = \prod_j P(D_j|H_iX)$$

and

$$P(D_1...D_m|\overline{H}_iX) = \prod_j P(D_j|\overline{H}_iX)$$

for $1 \leq i \leq n$ and $n > 2$ show that for any fixed i at most one of

$$\frac{P(D_j|H_iX)}{P(D_j|\overline{H}_iX)}$$

is not equal to 1.

Proof:

Firstly, we claim that the case of 2 pieces of data implies the general result.

Indeed, independence assumptions of all D_j together imply analogous pairwise independence for any pair D_k and D_l, and so, assuming the case with two data pieces is solved, for any pair $\frac{P(D_k|H_iX)}{P(D_k|\overline{H}_iX)}$, $\frac{P(D_l|H_iX)}{P(D_l|\overline{H}_iX)}$ at most one is not equal to 1, so of the whole set of $\frac{P(D_j|H_iX)}{P(D_j|\overline{H}_iX)}$ at most one is not equal to 1.

For a given hypothesis H_i we let $P(D_1|H_iX) = a_i$ $P(D_2|H_iX) = b_i$, so the ditribution of D_1 under H_iX is given by a vector $v_i = \begin{bmatrix} a_i \\ 1-a_i \end{bmatrix}$ and that of D_2 by $u_i = \begin{bmatrix} b_i \\ 1-b_i \end{bmatrix}$

Then independence of D_1 and D_2 says that the joint distribution of D_1D_2 (conditional on H_iX) is a product of distributions of D_1 and D_2 and is given by matrix

$$v_iu_i^T = \begin{bmatrix} a_i \\ 1-a_i \end{bmatrix} \begin{bmatrix} b_i & 1-b_i \end{bmatrix} =$$

$$= \begin{bmatrix} a_ib_i & a_i(1-b_i) \\ b_i(1-a_i) & (1-a_i)(1-b_i) \end{bmatrix}$$

Then the joint probability matrix of $D_1 D_2$ conditional on $\overline{H_i}$ is obtained by taking all the matrices of H_j with $j \neq i$ weighing them by (prior) probabilities h_j of H_j and adding them (and then dividing by the sum of the weights, but this is an overall normalizing factor which will not be important for us). That is, the matrix is proportional to (all sums are over $j \neq i$)

$$\sum_j h_j v_j u_j^T = \begin{bmatrix} \sum_j h_j a_j b_j & \sum_j h_j a_j (1-b_j) \\ \sum_j h_j b_j (1-a_j) & \sum_j h_j (1-a_j)(1-b_j) \end{bmatrix}$$

Now the assumption that D_1 and D_2 are independent conditional on $\overline{H}_i$ means this matrix is also a product of marginal distributions of $D_1|\overline{H}_i X$ and $D_2|\overline{H}_i X$, i.e. is of rank 1. This measn that it has determiant 0.

Let's start with the case of just 3 hypothesis. Start with $i = 3$.

Then a "conceptual" proof is as follows:

######Lemma: A sum M of two rank 1 matrices $M = h_1 v_1 u_1^T + h_2 v_2 u_2^T$ can only be rank 1 if either v_1 and v_2 are linearly dependent or u_1 and u_2 are linearly dependent.

Indeed, consider the image of M. $M(u_1)$ and $M(u_2)$ are both linear combinations of $h_1 v_1$ and $h_2 v_2$, so, to get that the image of M is span of v_1 and v_2, it is enough that the matrix of coefficients $G = \begin{pmatrix} u_1^T u_1 & u_1^T \cdot u_2 \\ u_2^T u_1 & u_2^T \cdot u_2 \end{pmatrix}$ is invertible. But G is the Grammian of u_1, u_2 and is invertible precisely when u_1 and u_2 are linearly independent (its determinant is the square of the area of the parallelogram spanned by u_1 and u_2, as you can easily verify). In that case (of independent us), the rank of M is the dimension of the span of v_1, v_2 and is 2, not 1 if v_1, v_2 are independent, proving what we want.

Remark: Those familiar with tensors may realize that we use metric in which u_1 and u_2 are orthonormal (that's the inverse of G) to "raise and index" and go from a bilinear form encoded by M to a linear map, whose range is then the span of vs.

Now if v_1 is linearly dependent with v_2 given that they are both probability vectors, this means $v_1 = v_2$, and similarly for us. So, either $v_1 = v_2$ or $u_1 = u_2$.

Now from $i = 1$ and $i = 2$ we get that (either $v_2 = v_3$ or $u_2 = u_3$) and (either $v_1 = v_3$ or $u_1 = u_3$). Since we have 3 equalities and only two types of vectors, either vs are equal twice, and $v_1 = v_2 = v_3$, or us are (and $u_1 = u_2 = u_3$). Correspondingly either D_1 has same distribution under all 3 hypothesis, and then $\frac{P(D_1|H_i X)}{P(D_1|\overline{H}_i X)}$ are all equal to 1, or D_2 does (and then all $\frac{P(D_1|H_i X)}{P(D_1|\overline{H}_i X)}$ are equal to 1). In either case, we get what we want.

Alternatively, for those who don't like linear algebra computational proof is as follows:

A (non-zero) 2 by 2 matrix has rank one when its determinant is zero. Writing this out in our case we get:

$$[h_1a_1b_1 + h_2a_2b_2][h_1(1-a_1)(1-b_1) + h_2(1-a_2)(1-b_2)]$$

$$= [h_1a_1(1-b_1) + h_2a_2(1-b_2)][h_1(1-a_1)b_1 + h_2(1-a_2)b_2]$$

Additively canceling $h_1^2a_1(1-a_1)b_1(1-b_1)$ and $h_2^2a_2(1-a_2)b_2(1-b_2)$ and then dividing by $h_1h_2(1-a_1)(1-a_2)(1-b_1)(1-b_2)$ and denoting $A_i = \frac{a_i}{1-a_i}$ and $B_i = \frac{b_i}{1-b_i}$ we get

$$A_1B_1 + A_2B_2 = A_1B_2 + A_2B_1$$

$$(A_1 - A_2)(B_1 - B_2) = 0$$

so either $A_1 = A_2$ or $B_1 = B_2$. Observe that $A_1 = A_2$ means $a_1 = a_2$ (equal odds means equal probability).

From here on it's the same: using $i = 2$ we get either $A_1 = A_3$ or $B_1 = B_3$ and $i = 1$ we get either $A_2 = A_3$ or $B_2 = B_3$. Since there are 2 choices for A or B and 3 times this choice is made, we will either have $A_1 = A_2 = A_3$ or $B_1 = B_2 = B_3$. In the first case D_1 is equally likely under all 3 hypothesis, and so $\frac{P(D_1|H_iX)}{P(D_1|\overline{H}_iX)}$ are all equal to 1. In the other case all $\frac{P(D_2|H_iX)}{P(D_2|\overline{H}_iX)}$ are equal to \$1

To get the extension to more than 3 hypothesis we use the following approach. As we mentioned before, a 2 by 2 matrix is or rank at most 1 if its determinant is zero. So we need some an efficient way of telling when the deterimnant of 2 by 2 matrix is zero.

Remark: More generally, a matrix is of rank at most 1 if all 2 by 2 minors have determinannt zero i.e. all $M^{\wedge 2}_{(i,j),(k,l)} = M_{ik}M_{jl} - M_{il}M_{jk}$ are zero. In tensor analysis, these are the entries of the second exterior power $M^{\wedge 2}$ of M. When dimension is 2 there is only one minor, and the $M^{\wedge 2}$ is a scalar, equal to $\det M$. So in dimensions above 2, we can reformulate everything below in these terms.

We will use the following property of 2D matrices M and N:

$$D(M,N) := \frac{1}{2}(\det(M+N) - \det M - \det N)$$

is symmetric and bilinear in M, N. This means

$$D(M,N) = D(N,M)$$

and

$$D(M_1 + M_2, N) = D(M_1, N) + D(M_1, N)$$

(and hence the same for second variable). Indeed, one computes

$D(M, N) = M_{11}N_{22} + M_{22}N_{11} - M_{12}N_{21} - M_{21}N_{12}$

and the resulting formula is linear in M and in N, i.e. bilinear.

Observe that $D(M, M) = \det M$. We then have, by induction on the number of summands,

$\det(\sum M_i) = D(\sum M_i, \sum M_j) = \sum_{i,j} D(M_i, M_j)$

Remark: We also have $D(\lambda M, N) = \lambda D(M, N)$, as usual in bilinearity, but we don't need this. In higher dimensions, and using tensor language, we are saying that taking second exterior power, which is quadratic, is a restriction of a symmetric bilinear operation.

Now we can apply this to our problem. Let $M_i = h_i v_i u_i^T$ and $N_i = \sum_{i \neq j} M_i$, and $M = M_i + N_i = \sum_j M_j$.

Our assumptions are that all M_i and N_j are rank 1 (i.e. have zero determinant). We now show that M has rank 1 (i.e. has zero determinant).

To that end we write

$$\det M = \sum_{j,k} D(M_j, M_k)$$

We want to see that this is zero. We know

$$0 = \det(N_i) = \sum_{j \neq i, k \neq i} D(M_j, M_k)$$

Summing over i we get (taking note that each $D(M_l, M_l)$ will apear $n-1$ times, while those $D(M_j, M_k)$ with $j \neq k$ will appear only $n-2$ times):

$$\sum_l D(M_l, M_l)(n-2) \sum_{j,k} D(M_j, M_k) = 0$$

So, since $D(M_l, M_l) = 0$, as long as $n \neq 2$ we have what we want.

This gives $M = vu^T$. Going back to $M = M_i + N_i$ we again see two rank one matrices add up to a rank one matrix. We conclude that for each i either $v = v_i$ (D_1 or is as likely under $H_i X$ as it is just under X), or $u = u_i$.

(TO DO: this leads either to all u_i equal or all v_i equal).