

## Elementary hypothesis testing

← Back to Chapters

### Exercise 4.1

Given

$$P(D_1 \dots D_m | H_i X) = \prod_j P(D_j | H_i X)$$

and

$$P(D_1 \dots D_m | \overline{H}_i X) = \prod_j P(D_j | \overline{H}_i X)$$

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

$$\frac{P(D_j | H_i X)}{P(D_j | \overline{H}_i X)}$$

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 a fixed  $i$  and for any pair of  $k, l$   $\frac{P(D_k | H_i X)}{P(D_k | \overline{H}_i X)}, \frac{P(D_l | H_i X)}{P(D_l | \overline{H}_i X)}$  at most one is not equal to 1, so of the whole set of  $\frac{P(D_j | H_i X)}{P(D_j | \overline{H}_i X)}$  at most one is not equal to 1.

So we focus on the case of only two data sets,  $D_1$  and  $D_2$ .

We will denote probability density/mass function of  $D_1$  under hypothesis  $H_i X$  by  $V_i$  and that of  $D_2$  by  $U_i$ . We will also denote probability density/mass function of  $D_1$  under hypothesis  $\overline{H}_i X$  by  $V_i^c$  and that of  $D_2$  by  $U_i^c$ , though we will not use these until the very end.

Remark 1: The proof actually works for arbitrary (not necessarily discrete or continuous) real-valued random variables, one just has to say that  $V$ s and  $U$ s are CDFs instead of PMFs/PDFs.

We will denote value of  $V_i$  at some  $D_1 = x_1$  by  $v_{i1}$  (and at  $D_1 = x_2$  by  $v_{i2}$ ). Similarly the values of  $U$  at  $D_2 = y_1$  will be denoted by  $u_{i1}$ .

We will use independence of  $D_1$  and  $D_2$  conditional on various hypotheses to prove independence under other hypotheses. Independence always means

$P(D_1 = x, D_2 = y|H) = P(D_1 = x|H)P(D_2 = y|H)$  and can be checked by checking for specific values.

With this in mind, we pick any pair of possible value pairs  $D_1 = x_1, D_2 = y_1$  and  $D_1 = x_2, D_2 = y_2$ , fixed from now on, and form vectors  $v_i = \begin{bmatrix} v_{i1} \\ v_{i2} \end{bmatrix}$  and  $u_i = \begin{bmatrix} u_{i1} \\ u_{i2} \end{bmatrix}$

Then independence of  $D_1$  and  $D_2$  (conditional on  $H_i X$ ) says that the joint distribution of  $P(D_1 = x, D_2 = y|H_i X)$  is a product of distributions of  $D_1$  and  $D_2$  and is given (at  $x = x_1, x_2$  and  $y = y_1, y_2$ ) by the matrix

$$\begin{aligned} M_i &= v_i u_i^T = \begin{bmatrix} v_{i1} \\ v_{i2} \end{bmatrix} \begin{bmatrix} u_{i1} & u_{i2} \end{bmatrix} = \\ &= \begin{bmatrix} v_{i1}u_{i1} & v_{i1}u_{i2} \\ v_{i2}u_{i1} & v_{i2}u_{i2} \end{bmatrix} \end{aligned}$$

It follows from this that the (joint) probability matrix of  $D_1 D_2$  (again, at  $x = x_1, x_2$  and  $y = y_1, y_2$ ) conditional on  $\overline{H}_i X$  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

$$\begin{aligned} \sum_{j \neq i} h_j M_i &= \sum_{j \neq i} h_j v_j u_j^T = \\ &= \begin{bmatrix} \sum_{j \neq i} h_j v_{j1} u_{j1} & \sum_{j \neq i} h_j v_{j1} u_{j2} \\ \sum_{j \neq i} h_j v_{j2} u_{j1} & \sum_{j \neq i} h_j v_{j2} u_{j2} \end{bmatrix} \end{aligned}$$

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” matrices of  $D_1|\overline{H}_i X$  and  $D_2|\overline{H}_i X$ , i.e. is of rank 1. This means that it has determinant 0.

**Three hypothesis.** Let’s start with the case of only 3 hypothesis  $H_1, H_2, H_3$ . Start with  $i = 3$ .

##### 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.

A “conceptual” proof is as follows: Consider the image of  $M$ . Vectors  $M(u_1)$  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 the span of  $v_1$  and  $v_2$ , it is enough that the matrix of coefficients

$G = \begin{pmatrix} u_1^T u_1 & u_1^T u_2 \\ u_2^T u_1 & u_2^T u_2 \end{pmatrix}$  is invertible. But  $G$  is the Gramian 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  $u$ s), the rank of  $M$  is the dimension of the span of  $v_1, v_2$  and if  $v_1, v_2$  were independent, it would be 2. So if rank of  $M$  is below 2, then either  $u$ s or  $v$ s are dependent, as wanted.

Remark 2: 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  $v$ s.

Remark 3: Alternatively, for those who don't like linear algebra, a computational proof of Lemma 1 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:

$$\begin{aligned} & [h_1 v_{11} u_{11} + h_2 v_{21} u_{21}] [h_1 v_{12} u_{12} + h_2 v_{22} u_{22}] = \\ & [h_1 v_{11} u_{12} + h_2 v_{21} u_{22}] [h_1 v_{12} u_{11} + h_2 v_{22} u_{21}] \end{aligned}$$

Additively canceling  $h_1^2 v_{11} v_{12} u_{11} u_{12}$  and  $h_2^2 v_{21} v_{22} u_{21} u_{22}$  and then dividing by  $h_1 h_2$  we have

$$\begin{aligned} & v_{11} u_{11} v_{22} u_{22} + v_{21} u_{21} v_{12} u_{12} = \\ & v_{11} u_{12} v_{22} u_{21} + v_{21} u_{22} v_{12} u_{11} \end{aligned}$$

or

$$(v_{11} v_{22} - v_{12} v_{21})(u_{11} u_{22} - u_{12} u_{21}) = 0,$$

so either  $v_1$  and  $v_2$  are linearly dependent, or  $u_1$  and  $u_2$  are.

Continuing with the case of three hypothesis, recall  $v_i$  and  $u_i$  were likelihood/probability vectors of  $D_1$  taking values  $x_1, x_2$  and  $D_2$  taking values  $y_1, y_2$  (under hypothesis  $H_i X$ ).

Observe that a pair of non-zero functions  $V_1$  and  $V_2$  such that  $(V_1(x_1), V_1(x_2))$  is always proportional to  $(V_2(x_1), V_2(x_2))$  are “globally” proportional meaning  $V_1 = k V_2$  (take any  $x$  with  $V_2(x) \neq 0$  and make  $k = V_1(x)/V_2(x)$ ).

If distributions of  $D_1$  under  $H_1 X$  and  $H_2 X$  are different, then they are also not proportional. By the previous paragraph, this implies that there will be two values  $x_1$  and  $x_2$  giving unproportional  $v_1$  and  $v_2$ . Then for arbitrary pair of values  $y_1, y_2$  of  $D_2$  the corresponding vectors  $u_1$  and  $u_2$  are proportional, so,

again, by the previous paragraph, the whole probability mass/density functions  $U_1$  and  $U_2$  of  $D_2$  under  $H_1X$  and  $H_2X$  are proportional, ergo equal.

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 distributions ( $U$  and  $V$ ), either the  $V$ s are equal twice, and  $V_1 = V_2 = V_3$ , or the  $U$ s are (and  $U_1 = U_2 = U_3$ ). Correspondingly either  $D_1$  has same distribution under all 3 hypothesis, and then  $\frac{P(D_1|H_iX)}{P(D_1|\bar{H}_iX)}$  are all equal to 1, or  $D_2$  does (and then all  $\frac{P(D_2|H_iX)}{P(D_2|\bar{H}_iX)}$  are equal to 1). In either case, we get what we want.

This completes the case of 3 hypothesis.

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

Remark 4: More generally, a matrix is of rank at most 1 if all 2 by 2 minors have determinant zero i.e. all  $M_{(i,j),(k,l)}^{\wedge 2} = M_{ik}M_{jl} - M_{il}M_{jk}$  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 formulate everything that follows in terms of determinants.

We will use the following property of 2D determinants. If  $M$  and  $N$  are 2 by 2 matrices then

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

is symmetric and bilinear in  $M, N$ . This means

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

and

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

(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 5: We also have  $D(\lambda M, N) = \lambda D(M, N)$ , as usual in bilinearity, but we don't need this.

Remark 6: In higher dimensions, and using tensor language, we are saying that taking second exterior power, which is quadratic in the input, is a restriction of a symmetric bilinear operation (on two inputs).

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 appear  $n - 1$  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, just as in the case of 3 hypothesis, that for each specific  $i$ , either  $V_i^c = V_i$  and hence  $\frac{P(D_1|H_i X)}{P(D_1|\overline{H_i X})} = 1$

OR

$U_i^c = U_i$  and hence  $\frac{P(D_2|H_i X)}{P(D_2|\overline{H_i X})} = 1$ . This is exactly what we wanted to prove.