Chapter 3

The Bhattacharyya Basis

In this chapter the structure of the Bhattacharyya functions will be investigated. Some very simple but extremely powerful identities will be described that are of interest in their own right. These will prove invaluable in the asymptotic analysis of the projective scores.

3.1 Some fundamental identities

The first of these results provides a recurrence formula for calculating higher order Bhattacharyya basis elements in terms of the derivatives of lower order elements. Throughout this chapter the regularity assumptions introduced in Chapter 2, namely R(t) will be assumed. Also, all expectations are taken at the model, θ_0, unless otherwise stated.

Theorem 3.1

$$V_{i_1 i_2 \cdots i_t i_{t+1}} = \frac{\partial}{\partial \lambda^{i_{t+1}}} V_{i_1 i_2 \cdots i_t} + V_{i_1 i_2 \cdots i_t} V_{i_{t+1}}. \tag{3.1}$$

<u>**Proof**</u>

Denoting the joint density of the independent random variables by p, from the definition of $V_{i_1 i_2 \cdots i_t}$ we have that

$$p\, V_{i_1 i_2 \cdots i_t} = \frac{\partial^t}{\partial \lambda^{i_1} \partial \lambda^{i_2}, \cdots, \partial \lambda^{i_t}} p.$$

On differentiating both sides with respect to $\lambda^{i_{t+1}}$ we obtain

$$\frac{\partial}{\partial \lambda^{i_{t+1}}} p\, V_{i_1 i_2 \cdots i_t} + p \frac{\partial}{\partial \lambda^{i_{t+1}}} V_{i_1 i_2 \cdots i_t} = \frac{\partial^{t+1}}{\partial \lambda^{i_1} \partial \lambda^{i_2}, \cdots, \partial \lambda^{i_t} \lambda^{i_{t+1}}} p.$$

Dividing both sides through by p gives the result. ∎

As an example of the use of the formula, first consider the simplest case, that is when there is a single nuisance parameter λ, then (3.1) becomes $V_{11} = \frac{\partial}{\partial\lambda^1}V_1 + V_1^2$. In the case of two nuisance parameters, $\lambda = (\lambda^1, \lambda^2)$ then one of the identities that (3.1) yields is $V_{12} = \frac{\partial}{\partial\lambda^2}V_1 + V_1V_2$.

The recurrence formula proves very useful when calculating the Bhattacharyya functions of high order. This is made possible though the use of a symbolic computational software package. For the work in this thesis `Mathematica` Wolfram (1991) was used extensively. To illustrate the use of the recurrence formula, consider the paired Binomial problem, Example 5.

For a fixed i, that is concentrating on a single strata, once the density has been defined all the code that is needed in `Mathematica` to calculate an arbitrary element of the Bhattacharyya Basis is

```
V[psi_,lambda_,t_] := V[psi,lambda,t] =
                         Simplify[D[V[psi,lambda,t-1],lambda]] +
                         Simplify[V[psi,lambda,t-1] V[psi,lambda,1]]
```

It is fair to say that the majority of the examples presented in this thesis would have been computationally unbearable without the aid of a symbolic software package.

The second result concerns certain expectation identities associated with functions in $\mathcal{G}_0$ that are orthogonal to the Bhattacharyya basis of a particular order. To be precise, let $T \in \mathcal{G}_0$ be a function that is orthogonal to the Bhattacharyya basis of order $t-1$, that is $T \perp \{\mathbf{V}_1, \mathbf{V}_2, \cdots \mathbf{V}_{t-1}\}$. Denote the derivative of T with respect to $\lambda^{i_1}, \lambda^{i_2} \cdots \lambda^{i_r}$ by $T_{i_1 i_2 \cdots i_r}$, and define for the case $r = 0$, $T_{i_0} = T$. Likewise, the derivatives of p with respect to $\lambda^{i_1}, \lambda^{i_2} \cdots \lambda^{i_r}$ will be denoted by $p_{i_1 i_2 \cdots i_r}$. Furthermore, the function V_{i_0} will be defined to be the constant function 1. Then

Theorem 3.2

$$E[T_{i_1 i_2 \cdots i_r} V_{i_{r+1} i_{r+2} \cdots i_t}] = (-1)^{t-r} E[T_{i_1 i_2 \cdots i_t}] \qquad 0 \le r \le t, \qquad y = 0, 1, \cdots .$$

From a symbolic perspective, the theorem states that the subscripts can be exchanged between the T and V functions, with the sign of the result alternating between -1 and $+1$ according to whether an odd or even number of subscripts have been exchanged.

Proof

The proof is by induction on t. The case where $t = 0$ is trivial, and the case when $t = 1$ is equivalent to the statement in (2.1). The first interesting case is $t = 2$ when the theorem asserts the following three identities:

$$\begin{aligned} E[T_{i_1 i_2} V_{i_0}] &= E[T_{i_1 i_2}]. \\ E[T_{i_1} V_{i_2}] &= -1\, E[T_{i_1 i_2}]. \\ E[T V_{i_1 i_2}] &= E[T_{i_1 i_2}]. \end{aligned}$$

The first of these is trivially true. The second is obtained from differentiating the identity $\int T V_{i_1} p dx = 0$ once with respect to λ^{i_2}, and the third arises from differentiating the identity $\int T p dx = 0$ twice with respect to λ^{i_1} and λ^{i_2} and using the orthogonality of T and V_{i_1}.

Now assume that the result is true for $T \perp \{\mathbf{V}_1, \mathbf{V}_2, \cdots \mathbf{V}_{t-2}\}$. Then

$$\begin{aligned} E[T_{i_1 i_2 \cdots i_r} V_{i_{r+1} i_{r+2} \cdots i_t}] &= \int T_{i_1 i_2 \cdots i_r} p_{i_{r+1} i_{r+2} \cdots i_t} dx \\ &= \int \left(T_{i_1 i_2 \cdots i_r} p_{i_{r+1} i_{r+2} \cdots i_t} + T_{i_1 i_2 \cdots i_{r+1}} p_{i_{r+2} \cdots i_t} - T_{i_1 i_2 \cdots i_{r+1}} p_{i_{r+2} \cdots i_t} \right) dx \\ &= \int \frac{\partial}{\partial \lambda^{i_{r+1}}} \left(T_{i_1 i_2 \cdots i_r} p_{i_{r+2} \cdots i_t} \right) dx - \int T_{i_1 i_2 \cdots i_{r+1}} p_{i_{r+2} \cdots i_t} dx \\ &= \frac{\partial}{\partial \lambda^{i_{r+1}}} E[T_{i_1 i_2 \cdots i_r} V_{i_{r+2} \cdots i_t}] - E[T_{i_1 i_2 \cdots i_{r+1}} V_{i_{r+2} \cdots i_t}]. \end{aligned}$$

But by the induction hypothesis $E[T_{i_1i_2\cdots i_r}V_{i_{r+2}\cdots i_t}] = (-1)^{t-(r+1)}E[T_{i_1i_2\cdots i_r i_{r+2}\cdots i_t}] = (-1)^r E[TV_{i_1i_2\cdots i_r i_{r+2}\cdots i_t}] = 0$ from the orthogonality assumption. It follows that

$$E[T_{i_1i_2\cdots i_r}V_{i_{r+1}i_{r+2}\cdots i_t}] = -E[T_{i_1i_2\cdots i_{r+1}}V_{i_{r+2}\cdots i_t}].$$

The same argument is repeated until all the subscripts have been transferred from V to T, which provides the result. ■

Corollary 3.1 *If* $T \perp \{\mathbf{V}_1, \mathbf{V}_2 \cdots \mathbf{V}_{t-1}\}$ *then*

$$E[T_{i_1i_2\cdots i_r}] = 0, \ \textit{for}\ r = 0, 1, \cdots t-1.$$

Proof

If $T \perp \{\mathbf{V}_1, \mathbf{V}_2 \cdots \mathbf{V}_{t-1}\}$, then $T \perp \{\mathbf{V}_1, \mathbf{V}_2 \cdots \mathbf{V}_r\}$ for any $r \leq t-1$. Applying Theorem 3.2 to $E[T_{i_1i_2\cdots i_r}]$ implies that

$$E[T_{i_1i_2\cdots i_r}] = (-1)^r E[TV_{i_1i_2\cdots i_r}],$$

but since $r \leq t-1$ and $T \perp \{\mathbf{V}_1, \mathbf{V}_2 \cdots \mathbf{V}_{t-1}\}$, $E[TV_{i_1i_2\cdots i_r}]$ must equal 0 which gives the result. ■

3.2 The extended information matrix

The projective scores U_t are conveniently expressed using the *Extended Information Matrix* M_t. Denoting the score function for the parameters of interest ψ by U_0, and the concatenated vector of the first t Bhattacharyya basis functions $\mathbf{V}_1, \mathbf{V}_2 \cdots \mathbf{V}_t$ by $\mathbf{V}$, then the *extended information matrix of order* t is defined as

$$M_t = E\left[\begin{pmatrix} U_0 \\ \mathbf{V} \end{pmatrix} \left(U_0^T, \mathbf{V}^T\right)\right].$$

The name "extended information matrix" is well deserved since M_1 is the usual Fisher's information matrix. Partitioning M_t as

$$M_t = \begin{pmatrix} E[U_0 U_0^T] & E[U_0 \mathbf{V}^T] \\ E[\mathbf{V} U_0^T] & E[\mathbf{V}\mathbf{V}^T] \end{pmatrix} = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix},$$

then the t^{th} projective score U_t can be written as

$$U_t = U_0 - M_{12} M_{22}^{+} \mathbf{V}, \tag{3.2}$$

and its associated information matrix is

$$I_{U_t} = M_{11} - M_{12} M_{22}^{+} M_{21}. \tag{3.3}$$

Clearly the matrix M_{22} plays a fundamental role, and it is of interest to examine its explicit form. This will be done in the context of an independent identically distributed sampling regime, and to illustrate the notation the case with two nuisance parameters, $\lambda = (\lambda^1, \lambda^2)$ will be used. Finally, the calculations will be carried out for the M_2 extended information matrix.

The probability density of the n $i.i.d.$random variables is as usual denoted by p, which by independence factors as $p = \prod_{i=1}^{n} p_i$,; the identically distributed assumption states that $p_i = p_j$ for all i and j. The aim is to compare the structure of the matrix M_{22} based on a sample of size 1 with the matrix M_{22} based on a sample of size n. Associating the superscript k with the k^{th} random variable, the functions in the Bhattacharyya basis of order 2 for the density p_k are $\left\{V_1^k, V_2^k, V_{11}^k, V_{12}^k, V_{21}^k, V_{22}^k\right\}$.

The matrix M_{22} based on a sample of size 1 is

$$M_{22} = \begin{pmatrix} E[V_1^kV_1^k] & E[V_1^kV_2^k] & E[V_1^kV_{11}^k] & E[V_1^kV_{12}^k] & E[V_1^kV_{21}^k] & E[V_1^kV_{22}^k] \\ E[V_2^kV_1^k] & E[V_2^kV_2^k] & E[V_2^kV_{11}^k] & E[V_2^kV_{12}^k] & E[V_2^kV_{21}^k] & E[V_2^kV_{22}^k] \\ E[V_{11}^kV_1^k] & E[V_{11}^kV_2^k] & E[V_{11}^kV_{11}^k] & E[V_{11}^kV_{12}^k] & E[V_{11}^kV_{21}^k] & E[V_{11}^kV_{22}^k] \\ E[V_{12}^kV_1^k] & E[V_{12}^kV_2^k] & E[V_{12}^kV_{11}^k] & E[V_{12}^kV_{12}^k] & E[V_{12}^kV_{21}^k] & E[V_{12}^kV_{22}^k] \\ E[V_{21}^kV_1^k] & E[V_{21}^kV_2^k] & E[V_{21}^kV_{11}^k] & E[V_{21}^kV_{12}^k] & E[V_{21}^kV_{21}^k] & E[V_{21}^kV_{22}^k] \\ E[V_{22}^kV_1^k] & E[V_{22}^kV_2^k] & E[V_{22}^kV_{11}^k] & E[V_{22}^kV_{12}^k] & E[V_{22}^kV_{21}^k] & E[V_{22}^kV_{22}^k] \end{pmatrix} \tag{3.4}$$

The identically distributed assumption means that (3.4) does not depend on k, so using an obvious notation it can be written as

$$M_{22} = \begin{pmatrix} W_{1,1} & W_{1,2} & W_{1,11} & W_{1,12} & W_{1,21} & W_{1,22} \\ W_{2,1} & W_{2,2} & W_{2,11} & W_{2,12} & W_{2,21} & W_{2,22} \\ W_{11,1} & W_{11,2} & W_{11,11} & W_{11,12} & W_{11,21} & W_{11,22} \\ W_{12,1} & W_{12,2} & W_{12,11} & W_{12,12} & W_{12,21} & W_{12,22} \\ W_{21,1} & W_{21,2} & W_{21,11} & W_{21,12} & W_{21,21} & W_{21,22} \\ W_{22,1} & W_{22,2} & W_{22,11} & W_{22,12} & W_{22,21} & W_{22,22} \end{pmatrix}. \tag{3.5}$$

The following lemma, which will be generalized in Section 3 of this chapter, expresses the elements of the Bhattacharyya basis for a sample of size n, in terms of the Bhattacharyya basis functions based on the sample of size 1. Note that the assumption that the random variables are identically distributed is not needed.

Lemma 3.1

$$V_i = \sum_{k=1}^{n} V_i^k. \tag{3.6}$$

$$V_{ij} = \sum_{k=1}^{n} V_{ij}^k + \sum_{k=1}^{n} \sum_{l \neq k}^{n} V_i^k V_j^l. \tag{3.7}$$

Proof

By definition

$$V_i = \frac{\frac{\partial}{\partial\lambda^i}\prod_{k=1}^n p_k}{\prod_{k=1}^n p_k} = \sum_{k=1}^n V_i^k.$$

Using the recurrence relation (Theorem 3.1) we have that

$$V_{ij} = \frac{\partial}{\partial\lambda^j} V_i + V_i V_j. \tag{3.8}$$

Substituting the expression for V_i given by (3.6) into (3.8) it follows that

$$V_{ij} = \sum_{k=1}^n \frac{\partial}{\partial\lambda^j} V_i^k + \sum_{k=1}^n \sum_{l=1}^n V_i^k V_j^l, \tag{3.9}$$

but from the recurrence relation (3.1), $\frac{\partial}{\partial\lambda^j} V_i^k = V_{ij}^k - V_i^k V_i^k$. Substituting this expression back into (3.9) and rearranging gives the result (3.7). ■

The next lemma expresses the covariances of the elements of the Bhattacharyya basis, based on the sample of size n, in terms of the covariances of the Bhattacharyya basis elements based on a sample of size 1.

Lemma 3.2

$$\begin{aligned}
E[V_i V_r] &= \sum_{k=1}^n E[V_i^k V_r^k]. \\
E[V_i V_{rs}] &= \sum_{k=1}^n E[V_i^k V_{rs}^k]. \\
E[V_{ij} V_{rs}] &= \sum_{k=1}^n E[V_{ij}^k V_{rs}^k] + \sum_{k=1}^n \sum_{l\neq k}^n E[V_i^k V_r^k] E[V_j^l V_s^l] + \sum_{k=1}^n \sum_{l\neq k}^n E[V_i^k V_s^k] E[V_j^l V_r^l].
\end{aligned}$$

These results are easily proved using the independence of the random variables and the fact that all the elements of the Bhattacharyya basis are mean 0.

As a special case when the observations are identically distributed, we have

Corollary 3.2

$$E[V_i V_r] = nE[V_i^k V_r^k] = nW_{i,r}.$$

$$
\begin{aligned}
E[V_i V_{rs}] &= nE[V_i^k V_{rs}^k] = nW_{i,rs}. \\
E[V_{ij} V_{rs}] &= nE[V_{ij}^k V_{rs}^k] + n(n-1)E[V_i^k V_r^k]E[V_j^l V_s^l] + n(n-1)E[V_i^k V_s^k]E[V_j^l V_r^l], \\
&= nW_{ij,rs} + n(n-1)W_{i,r}W_{j,s} + n(n-1)W_{i,s}W_{j,r}. \qquad (3.10)
\end{aligned}
$$

Note that the expectation of $V_{ij}V_{rs}$ is $O(n^2)$ with leading coefficient $W_{i,r}W_{j,s} + W_{i,s}W_{j,r}$. This is an important structural identification, and shows that the matrix comprising elements $E[V_{ij}V_{rs}]$, in which the rows are indexed by (i, j) and the columns by (r, s) is dominated by an order n^2 matrix with elements $\{W_{i,r}W_{j,s} + W_{i,s}W_{j,r}\}$. Why is this so useful? To answer this question, note how the dimensionality of the problem is increasing. If there are s nuisance parameters, then $\mathbf{V}_i$ is a vector with s elements, and its covariance matrix is of dimension $s \times s$. There are s^2 terms of the form V_{ij}, so that the covariance matrix of V_{ij} is of dimensions $s^2 \times s^2$, and clearly the dimensionality of the problem is increasing as a polynomial in s. All these covariance matrices need to be inverted at some stage of the calculations, and this may pose severe computational problems. But, a matrix with elements

$$W_{i,r}W_{j,s} + W_{i,s}W_{j,r} \qquad (3.11)$$

possesses a very simple Moore-Penrose generalized inverse. It is simply

$$\tfrac{1}{4}\left(W^{i,r}W^{j,s} + W^{i,s}W^{j,r}\right) \qquad (3.12)$$

(see (McCullagh, 1987, p.20, Exercise 1.12)), where $W^{i,r}$ denotes the elements of the inverse of $W_{i,r}$. This means that the *leading terms*[1] in the inverse of the matrix with $(ij, rs)^{th}$ element $E[V_{ij}V_{rs}]$, can be calculated by inverting an $s \times s$ dimensional matrix, rather than an $s^2 \times s^2$ dimensional one.

To demonstrate the use of the extended information matrix, the Neyman-Scott problem of Example 2 based on one stratum will be used. Since there is a

[1] This is not an immediate statement, but will be justified in the last section of this chapter.

solitary stratum, there is no longer a need for the subscript i to denote the strata. For the Neyman-Scott problem the U_0 and V_1 scores are given by equations (1.1) and (1.2) respectively. The V_{11} function is from Theorem 3.1 equal to

$$\frac{\partial}{\partial\lambda}V_1 + (V_1)^2 = -\frac{n}{\psi} + \left(\frac{1}{\psi}\sum_{j=1}^{n}(x_j-\lambda)\right)^2.$$

The extended information matrix can be calculated as

$$M_2 = \begin{pmatrix} \frac{n}{2\psi^2} & 0 & -\frac{n}{\psi^2} \\ 0 & \frac{n}{\psi} & 0 \\ -\frac{n}{\psi^2} & 0 & -\frac{2\,n^2}{\psi^2} \end{pmatrix}.$$

Consequently, utilizing (3.2) the U_2 projective score is

$$U_2 = U_0 + \frac{1}{2\,n}V_{11},$$

which is in fact *identical* to the conditional score U_c. Here is another example where a low order projective score is identical to the conditional score. Furthermore, from (3.3) the information in U_2, that is I_{U_2}, is equal to

$$\frac{n}{2\psi^2} - \left(-\frac{n}{\psi^2}\right)\left(\frac{2\,n^2}{\psi^2}\right)^{-1}\left(-\frac{n}{\psi^2}\right) = \frac{1}{2}\left(\frac{n-1}{\psi^2}\right).$$

3.3 U-statistic structure of the Bhattacharyya basis elements

Since $V_{i_1 i_2 \cdots i_t}$ is by definition equal to

$$\frac{\frac{\partial}{\partial\lambda^{i_1}\partial\lambda^{i_2},\ldots,\partial\lambda^{i_t}}\prod_{i=1}^{n}p_i(x_i,\psi,\lambda)}{\prod_{i=1}^{n}p_i(x_i,\psi,\lambda)}, \tag{3.13}$$

the explicit determination of $V_{i_1 i_2 \cdots i_t}$ amounts to finding, and establishing a notation for the higher order derivatives of a product.

3.3.1 The lattice of set partitions

Finding the higher order derivatives involves determining all possible partitions of indices of the $\lambda^{i_1}, \lambda^{i_2}, \cdots, \lambda^{i_t}$, a task that is conveniently handled using notation from the theory of lattices. The basic ideas needed to understand the partition arguments are outlined in the following section (they are again taken from McCullagh (1987)):

By definition, a lattice $\mathcal{L}$, is a finite partially ordered set, such that every pair of elements $a, b \in \mathcal{L}$ has a unique greatest lower bound, and a unique least upper bound. Set partitions form a lattice with the order relation being defined by $a < b$ if a is a sub-partition of b.

In order to illustrate the notation and terminology consider partitioning the indices $i_1, i_2 \cdots i_6$. A partition is identified by the number of elements being partitioned, (in this example 6) and the number of blocks that constitute the partition. For example, $\{i_1 i_2 i_3 | i_4 i_5 i_6\}$ represents a partition comprising 2 blocks. The number of blocks is referred to as the order of the partition and the number of elements as the degree, so that the above example is a partition of degree 6 and order 2. Two special cases are when the order of the partition, say α, is equal to the degree, say t, and when the order of the partition is equal to 1. These cases correspond to the partitions $\{i_1 | i_2 | i_3 | i_4 | i_5 | i_6\}$ and $\{i_1 i_2 i_3 i_4 i_5 i_6\}$, and are denoted by Υ_t and Υ_1 respectively. An arbitrary partition Υ of t indices into α blocks will be denoted by $\{v_1, v_2 \cdots v_\alpha\}$, where each v_i contains a subset of the t indices.

Using this notation $V_{i_1 i_2 \cdots i_t}$ has the following form:

Proposition 3.1

$$V_{i_1 i_2 \cdots i_t} = \sum_{\Upsilon_t \leq \Upsilon \leq \Upsilon_1} g(\Upsilon),$$

where

$$g(\Upsilon) = \sum_{r_1 \neq r_2 \cdots \neq r_\alpha} V_{\upsilon_1}^{r_1} V_{\upsilon_2}^{r_2} \cdots V_{\upsilon_\alpha}^{r_\alpha} [\tau(\Upsilon)], \tag{3.14}$$

and $\tau(\Upsilon)$ is the number of non-distinct partitions of Υ, that is, partitions that are equivalent up to identification of the indices.

The bracket [ ] notation is used in order to avoid explicitly listing the non-distinct partitions, and the reader is referred to McCullagh (1987, p.29) for further details of the notation.

Proof

Differentiating the product $\prod_{i=1}^{n} p_i(x_i, \psi, \lambda)$ with respect to $\lambda^{i_1}, \lambda^{i_2}, \cdots, \lambda^{i_t}$ amounts to allocating the indices $\{i_1, i_2, \cdots, i_t\}$ to the integers $\{1, 2, \cdots, n\}$, then differentiating the p_i with respect to the appropriate λ^{i_j} and summing over all the possible allocations of indices to integers. The allocation procedure allows for the possibility that more than one index may be allocated to a specific integer, and that some integers will be allocated no indices. The allocation induces a natural partitioning of the indices, with a block being associated with each integer. Some blocks will contain no indices, indicating that the particular p_i is not differentiated at all. Consequently, it will cancel with itself in the denominator of (3.13). Let Υ denote a partition with α non empty blocks. This partition will be repeated on all groups of α distinct integers, so that the term in the derivative associated with it can be written as

$$\sum_{r_1 \neq r_2 \cdots \neq r_\alpha} V_{\upsilon_1}^{r_1} V_{\upsilon_2}^{r_2} \cdots V_{\upsilon_\alpha}^{r_\alpha}.$$

Grouping all partitions that are equivalent up to identification of the indices and summing over Υ gives the result. ■

To clarify the notation consider the case where $t = 3$ so that $V_{i_1 i_2 i_3}$ is of interest. Then the possible partitions of $\{i_1i_2i_3\}$ are $\Upsilon_1 = \{i_1i_2i_3\}$, $\{i_1i_2|i_3\}$ and $\Upsilon_3 = \{i_1|i_2|i_3\}$. It follows that

$$\begin{aligned} V_{i_1i_2i_3} = \sum_{r_1} V^{r_1}_{i_1i_2i_3}[\tau(\{i_1i_2i_3\})] \;&+\; \sum_{r_1\neq r_2} V^{r_1}_{i_1i_2}V^{r_2}_{i_3}[\tau(\{i_1i_2|i_3\})] \\ &+\; \sum_{r_1\neq r_2\neq r_3} V^{r_1}_{i_1}V^{r_2}_{i_2}V^{r_3}_{i_3}[\tau[\{i_1|i_2|i_3\})]. \end{aligned}$$

In this case

$$[\tau(\{i_1i_2i_3\})] = [1] \quad [\tau(\{i_1i_2|i_3\})] = [3] \quad [\tau[\{i_1|i_2|i_3\})] = [1],$$

since there is only one combination of the indices $\{i_1i_2i_3\}$ but three of the indices $\{i_1i_2|i_3\}$ namely $\{i_1i_2|i_3\}$,$\{i_1i_3|i_2\}$ and $\{i_2i_3|i_1\}$, and again there is only one combination of the indices $\{i_1|i_2|i_3\}$. This means that the term

$$\sum_{r_1\neq r_2} V^{r_1}_{i_1i_2}V^{r_2}_{i_3}[\tau(\{i_1i_2|i_3\})] = \sum_{r_1\neq r_2} V^{r_1}_{i_1i_2}V^{r_2}_{i_3}[3]$$

represents the sum

$$\sum_{r_1\neq r_2} V^{r_1}_{i_1i_2}V^{r_2}_{i_3} + \sum_{r_1\neq r_2} V^{r_1}_{i_1i_3}V^{r_2}_{i_2} + \sum_{r_1\neq r_2} V^{r_1}_{i_2i_3}V^{r_2}_{i_1}. \tag{3.15}$$

It is possible in some cases to simplify this term. In particular, if λ is univariate, so that $\dim(\lambda) = 1$ then $i_1 = i_2 = i_3 = 1$, then (3.15) reduces to $3\sum_{r_1\neq r_2} V^{r_1}_{11}V^{r_2}_{1}$. If $\dim(\lambda) = 2$, and $i_1 = i_2 = 1, i_3 = 2$ then it simplifies to

$$\sum_{r_1\neq r_2} V^{r_1}_{11}V^{r_2}_{2} + 2\sum_{r_1\neq r_2} V^{r_1}_{12}V^{r_2}_{1}.$$

The number of non-distinct partitions equivalent to $\Upsilon = \{\upsilon_1, \upsilon_2 \cdots \upsilon_\alpha\}$, in other words $\tau(\Upsilon)$ is given by the formula

$$\tau(\Upsilon) = \frac{t!}{a_1!a_2!\cdots a_\alpha!}\frac{1}{b_1!b_2!\cdots b_t!},$$

where α is the degree of the partition, a_i is the number of elements in the set v_i, and b_j is the number of the v_i with exactly j elements. The formula is just an application of the multinomial theorem, with the consideration that if the multinomial theorem is thought of as the number of ways that α distinct objects can be allocated into groups of size $a_1, a_2 \cdots a_t$, then in this set up groups of the same size are indistinguishable. To demonstrate, let $\alpha = 7$, and consider the partition $\{i_1 i_2 | i_3 i_4 | i_5 i_6 i_7\}$. There are 2 partitions of size 2 and one partition of size 3, consequently

$$\tau(\{i_1 i_2 | i_3 i_4 | i_5 i_6 i_7\}) = \frac{7!}{2!2!3!}\frac{1}{2!1!} = 105,$$

cf (McCullagh, 1987, Table 2, p.257). McCullagh has tables for partitions of up to 8 indices and the reader is referred to these if calculations are to be explicitly made.

3.3.2 Identification of the embedded U-statistics

It is natural to group the partitions into classes of the same order, so that

$$V_{i_1 i_2 \cdots i_t} = \sum_{1 \leq \alpha \leq t} \sum_{\Upsilon_\alpha} g(\Upsilon_\alpha) = \sum_{1 \leq \alpha \leq t} h(\alpha), \text{ where } h(\alpha) = \sum_{\Upsilon_\alpha} g(\Upsilon_\alpha),$$

and Υ_α represents an arbitrary partition of order α.

Proposition 3.2 *For $\alpha_1 \neq \alpha_2$, $E[h(\alpha_1)h(\alpha_2)] = 0$*

The result indicates that the classes are uncorrelated.

Proof

Without loss of generality assume that $\alpha_1 < \alpha_2$. Picking arbitrary terms from $h(\alpha_1)$ and $h(\alpha_2)$ corresponding to partitions $\{v_1, v_2, \cdots, v_{\alpha_1}\}$ and $\{\tilde{v}_1, \tilde{v}_2, \cdots, \tilde{v}_{\alpha_2}\}$ and forming their product gives

$$\sum_{r_1 \neq r_2 \cdots \neq r_{\alpha_1}} V_{v_1}^{r_1} V_{v_2}^{r_2} \cdots V_{v_{\alpha_1}}^{r_{\alpha_1}} \times \sum_{s_1 \neq s_2 \cdots \neq s_{\alpha_2}} V_{\tilde{v}_1}^{s_1} V_{\tilde{v}_2}^{s_2} \cdots V_{\tilde{v}_{\alpha_2}}^{s_{\alpha_2}}.$$

Note that each $V_{\upsilon_i}^{r_i}$ is mean 0 and independent of all other $V_{\upsilon_j}^{r_j}$ for $i \neq j$. Consequently the expectation can be written as

$$\alpha_1! E\left[\sum_{r_1 \neq r_2 \cdots \neq r_{\alpha_1}} V_{\upsilon_1}^{r_1} V_{\tilde{\upsilon}_1}^{r_1} V_{\upsilon_2}^{r_2} V_{\tilde{\upsilon}_2}^{r_2} \cdots V_{\upsilon_{\alpha_1}}^{r_{\alpha_1}} V_{\tilde{\upsilon}_{\alpha_1}}^{r_{\alpha_1}}\right] \times$$
$$E\left[\sum_{r_1 \neq \cdots \neq r_{\alpha_1} \neq s_{\alpha_1+1} \neq \cdots \neq s_{\alpha_2}} V_{\tilde{\upsilon}_{\alpha_1+1}}^{s_{\alpha_1+1}} \cdots V_{\tilde{\upsilon}_{\alpha_2}}^{s_{\alpha_2}}\right]$$

and the result follows from noting that the second expectation is zero. ■

The order of magnitude of the covariance between terms such as $V_{i_1 i_2 \cdots i_{t_1}}$ and $V_{j_1 j_2 \cdots j_{t_2}}$ plays a key role in the subsequent theory. A simple extension of the argument in Proposition 3.2 shows that

Proposition 3.3

$$E[V_{i_1 i_2 \cdots i_{t_1}} V_{j_1 j_2 \cdots j_{t_2}}] = O(n^{min(t_1, t_2)}),$$

a special case of which is $\mathrm{Var}[V_{i_1 i_2 \cdots i_{t_1}}] = O(n^{t_1})$.

Proof

This is seen by noting that only classes of equal order in $V_{i_1 i_2 \cdots i_{t_1}}$ and $V_{j_1 j_2 \cdots j_{t_2}}$ can contribute to the expectation $E[V_{i_1 i_2 \cdots i_{t_1}} V_{j_1 j_2 \cdots j_{t_2}}]$, and the largest contributing class is of order $min(t_1, t_2)$. ■

Not only are the classes uncorrelated, but each class is itself a U-statistic, as the following theorem proves.

Theorem 3.3 *Each class of order α in $V_{i_1 i_2 \cdots i_t}$, that is $h(\alpha)$, when standardized by $\binom{n}{\alpha}$ is a mean 0 U-statistic. Furthermore, it is uncorrelated with all other U-statistics from any class of order $\alpha_1 \neq \alpha$.*

Proof

By definition

$$\begin{aligned} h(\alpha) &= \sum_{\Upsilon_\alpha} g(\Upsilon_\alpha) \\ &= \sum_{\Upsilon_\alpha} \left[\sum_{r_1 \neq r_2 \cdots \neq r_\alpha} V_{\upsilon_1}^{r_1} V_{\upsilon_2}^{r_2} \cdots V_{\upsilon_\alpha}^{r_\alpha} [\tau(\Upsilon_\alpha)] \right] \end{aligned}$$

Now fix Υ_α and consider

$$\sum_{r_1 \neq r_2 \cdots \neq r_\alpha} V_{\upsilon_1}^{r_1} V_{\upsilon_2}^{r_2} \cdots V_{\upsilon_\alpha}^{r_\alpha} = \sum_{1 \leq r_1 < r_2 \cdots < r_\alpha \leq n} \left[\sum_p \left(V_{\upsilon_1}^{r_1} V_{\upsilon_2}^{r_2} \cdots V_{\upsilon_\alpha}^{r_\alpha} \right) \right]$$

where $\sum_p$ denotes summation over the $\alpha!$ permutations of $\{r_1, r_2 \cdots r_\alpha\}$. The function $\sum_p \left(V_{\upsilon_1}^{r_1} V_{\upsilon_2}^{r_2} \cdots V_{\upsilon_\alpha}^{r_\alpha} \right)$ is a symmetric kernel of *i.i.d.* observations, so that

$$\sum_{1 \leq r_1 < r_2 \cdots < r_\alpha \leq n} \left[\sum_p \left(V_{\upsilon_1}^{r_1} V_{\upsilon_2}^{r_2} \cdots V_{\upsilon_\alpha}^{r_\alpha} \right) \right]$$

when divided by $\binom{n}{\alpha}$ is indeed a U-statistic. Furthermore it has mean equal to 0 since $E[V_{\upsilon_1}^{r_1} V_{\upsilon_2}^{r_2} \cdots V_{\upsilon_\alpha}^{r_\alpha}] = 0$. It follows that $h(\alpha)$ is a sum of order α U-statistics, and is therefore itself an order α U-statistic. The fact that it is uncorrelated with all other U-statistic of order $\alpha_1 \neq \alpha$ follows immediately from Proposition 3.2. $\blacksquare$

Serfling (1980) is a good source for information on U-statistics; the U-statistic results used in this thesis can be found in Chapter 5 of this book.

The next theorem demonstrates how $V_{i_1 i_2 \cdots i_t}$ is stochastically dominated by a single term, $g(\Upsilon_t)$.

Theorem 3.4

$$\frac{1}{n^{t/2}} \left(V_{i_1 i_2 \cdots i_t} - g(\Upsilon_t) \right) \xrightarrow{p} 0.$$

Proof

$$\frac{1}{n^{t/2}} \left(V_{i_1 i_2 \cdots i_t} - g(\Upsilon_t) \right) = \frac{1}{n^{t/2}} \sum_{\Upsilon_t < \Upsilon \leq \Upsilon_1} g(\Upsilon).$$

Now fix an Υ such that $\Upsilon_t < \Upsilon \leq \Upsilon_1$, then

$$\frac{1}{n^{t/2}}g(\Upsilon) = \frac{1}{n^{t/2}} \sum_{r_1 \neq r_2 \cdots \neq r_\alpha} V_{\upsilon_1}^{r_1} V_{\upsilon_2}^{r_2} \cdots V_{\upsilon_\alpha}^{r_\alpha}[\tau(\Upsilon)], \tag{3.16}$$

where $\alpha < t$. For each element $\sum_{r_1 \neq r_2 \cdots \neq r_\alpha} V_{\upsilon_1}^{r_1} V_{\upsilon_2}^{r_2} \cdots V_{\upsilon_\alpha}^{r_\alpha}$ of the RHS of (3.16) we have

$$\frac{1}{n^{t/2}} \sum_{r_1 \neq r_2 \cdots \neq r_\alpha} V_{\upsilon_1}^{r_1} V_{\upsilon_2}^{r_2} \cdots V_{\upsilon_\alpha}^{r_\alpha} = n^{\frac{\alpha - t}{2}} \sum_{r_1} \frac{V_{\upsilon_1}^{r_1}}{n^{\frac{1}{2}}} \sum_{r_2 \neq r_1} \frac{V_{\upsilon_2}^{r_2}}{n^{\frac{1}{2}}} \cdots \sum_{r_\alpha \neq r_{\alpha-1} \neq \cdots r_1} \frac{V_{\upsilon_\alpha}^{r_\alpha}}{n^{\frac{1}{2}}}. \tag{3.17}$$

Since $V_{\upsilon_j}^{r_j}$ are mean zero *i.i.d.* random variables the central limit theorem holds for each of the $\sum \frac{V_{\upsilon_j}^{r_j}}{n^{\frac{1}{2}}}$, $1 \leq j \leq \alpha$, that is $\sum \frac{V_{\upsilon_j}^{r_j}}{n^{\frac{1}{2}}} = O_p(1)$. The finite product of a series of $O_p(1)$ terms, is again $O_p(1)$, and since $\alpha < t$, $n^{\frac{\alpha - t}{2}} = o(1)$. Consequently the RHS of (3.17) is $o_p(1)$. It follows that $\frac{1}{n^{t/2}}\left(V_{i_1 i_2 \cdots i_t} - g(\Upsilon_t)\right)$ is the sum of a finite number of $o_p(1)$ terms, and is thus itself $o_p(1)$ which is the required result. ■

The dominating term in $V_{i_1 i_2 \cdots i_t}$ is $g(\Upsilon_t)$, that is the term corresponding to the complete partition of $\{i_1, i_2 \cdots i_t\}$ into $\alpha = t$ blocks, each with a single element. Since $\tau(\Upsilon_t) = 1$, the leading term in $V_{i_1 i_2 \cdots i_t}$ is $\sum_{r_1 \neq r_2 \cdots \neq r_t} V_{i_1}^{r_1} V_{i_2}^{r_2} \cdots V_{i_t}^{r_t}$.

3.4 Orthogonalization of the Bhattacharyya basis

The following results are derived in the *i.i.d.* sampling context. By use of the Gram-Schmidt procedure, an arbitrary set of functions may be orthogonalized and scaled so that they form an orthonormal basis. If the orthonormalized version of the Bhattacharyya basis is denoted by $\mathbf{V}^*$, then it has the property that $E[\mathbf{V}^* \mathbf{V}^{*T}] = I$, where I is an identity matrix. Clearly, the orthonormalized version of the basis will easier to work with from an analytic viewpoint. However, there are some advantages in not proceeding with the entire orthonormalization process, but to only take it as far as needed to elucidate the relevant structures. In particular, a partially orthogonalized basis will be used in which the Bhattacharyya function of order t are made orthogonal to all lower order Bhattacharyya functions, but are allowed to be correlated among themselves. This will give rise to an extended information matrix of the form

$$M_t = \begin{pmatrix} E[U_0 U_0^T] & E[U_0 \mathbf{V}_1^T] & E[U_0 \mathbf{V}_2^{*T}] & E[U_0 \mathbf{V}_2^{*T}] & \cdots & E[U_0 \mathbf{V}_t^{*T}] \\ E[V_1 U_0^T] & E[\mathbf{V}_1 \mathbf{V}_1^T] & 0 & 0 & \cdots & 0 \\ E[\mathbf{V}_2^* U_0^T] & 0 & E[\mathbf{V}_2^* \mathbf{V}_2^{*T}] & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \cdots & 0 \\ E[\mathbf{V}_t^* U_0^t] & 0 & 0 & \cdots & \cdots & E[\mathbf{V}_t^* \mathbf{V}_t^{*T}] \end{pmatrix}.$$

Note that without loss of generality we can set $\mathbf{V}_1^* = \mathbf{V}_1$.

In order to construct $\mathbf{V}_t^*$ all that is needed is to make $\mathbf{V}_t$ orthogonal to all lower order Bhattacharyya basis elements, that is a set of constants, $\{\mathbf{a}\}$, must be found for each $V_{i_1 i_2 \cdots i_t}$ such that

$$V^*_{i_1 i_2 \cdots i_t} = \tag{3.18}$$

$$V_{i_1 i_2 \cdots i_t} - \sum_{j_1, j_2 \cdots j_{t-1}} a_{j_1 j_2 \cdots j_{t-1}} V^*_{j_1 j_2 \cdots j_{t-1}} - \sum_{j_1, j_2 \cdots j_{t-2}} a_{j_1 j_2 \cdots j_{t-2}} V^*_{j_1 j_2 \cdots j_{t-2}} - \cdots - \sum_{j_1} a_{j_1} V_{j_1} \tag{3.19}$$

and $E[\mathbf{V}^*_{i_1 i_2 \cdots i_t} V_{i_1 i_2 \cdots i_j}] = 0$ for $j = 1, 2 \cdots t-1$.

The first claim is that these constants are all $O(1)$ so that $V^*_{i_1 i_2 \cdots i_t}$ is still dominated by the term $g(\Upsilon_t)$.

Lemma 3.3 *The constants* $\{\mathbf{a}\}$ *in (3.18) are* $O(1)$.

Proof

The proof is by induction on t. First examine the case $t = 2$, then

$$V^*_{i_1 i_2} = V_{i_1 i_2} - \sum_{j_1} a_{j_1} V_{j_1}, \tag{3.20}$$

and the a_{j_1} are given explicitly by the formula

$$a_{j_1} = E[V_{i_1 i_2} \mathbf{V}_1^T] E[\mathbf{V}_1 \mathbf{V}_1^T]^{-1}.$$

Recall that the matrix $E[\mathbf{V}_1 \mathbf{V}_1^T]$ is by assumption full rank. The elements of the vector $E[V_{i_1 i_2} \mathbf{V}_1]$ are all $O(n)$ by Proposition 3.3, and similarly all the elements of $E[\mathbf{V}_1 \mathbf{V}_1^T]$ are $O(n)$. Consequently the product $E[V_{i_1 i_2} \mathbf{V}_1] E[\mathbf{V}_1 \mathbf{V}_1^T]^{-1}$ is $O(1)$ as required. Clearly $V^*_{i_1 i_2}$ is still dominated by $g(\Upsilon_2)$.

Assume now that the result is true up to $t-1$ and consider a particular element, $\sum_{i_1, i_2 \cdots i_j} a_{i_1 i_2 \cdots i_j} V^*_{i_1 i_2 \cdots i_j}$ from (3.18). As before the constants $\{a_{i_1 i_2 \cdots i_j}\}$ are found from the equation

$$a_{i_1 i_2 \cdots i_j} = E[V_{i_1 i_2 \cdots i_t} \mathbf{V}_j^{*T}] E[\mathbf{V}_j^* \mathbf{V}_j^{*T}]^+,$$

and from the induction hypothesis all the functions in $\mathbf{V}_j^*$ are dominated by their respective $g(\Upsilon_j)$ so that $E[V_{i_1 i_2 \cdots i_t} \mathbf{V}_j^{*T}]$ has elements all of $O(n^j)$. Similarly $E[\mathbf{V}_j^* \mathbf{V}_j^{*T}]$ has elements of order $O(n^j)$. It follows[2] that the matrix $E[\mathbf{V}_j^* \mathbf{V}_j^{*T}]^+$ has terms of at most $O(n^{-j})$, and that $a_{i_1 i_2 \cdots i_j}$ is $O(1)$. Finally note that $V^*_{i_1 i_2 \cdots i_t}$ is still dominated by $g(\Upsilon_t)$ since the correction terms to $V_{i_1 i_2 \cdots i_t}$ contain at most $t-1$ subscripts and so are products of $t-1$ random variables, and are multiplied by a constant of order $O(1)$. ■

Since the $V_{i_1 i_2 \cdots i_t}$ are all dominated by their respective $g(\Upsilon_t)$, the matrix $E[\mathbf{V}_t^* \mathbf{V}_t^{*T}]$ has a dominating structure induced by the $g(\Upsilon_t)$:

Proposition 3.4 *The matrix $E[\mathbf{V}_t^* \mathbf{V}_t^{*T}]$, indexed by $\{i_1, i_2 \cdots i_t, j_1, j_2 \cdots j_t\}$ can be written as*

$$\frac{n!}{(n-t)!} \sum_p W_{i_1, j_1} W_{i_2, j_2} \cdots W_{i_t, j_t} + Rem,$$

[2]Again, this is not immediate, but see the final section of the chapter for an argument establishing this fact.

where $\sum_p$ indicates summation over all permutations of the subscripts $\{j_1, j_2 \cdots j_t\}$, and Rem is a matrix with elements of at most order $O(n^{t-1})$.

<u>Proof</u>

Consider an arbitrary term from the matrix $E[\mathbf{V}_t^* \mathbf{V}_t^{*T}]$, say

$$E[\sum_{r_1 \neq r_2 \cdots r_t} V^*_{i_1 i_2 \cdots i_t} \sum_{s_1 \neq s_2 \cdots s_t} V^*_{j_1 j_2 \cdots j_t}].$$

If the dominant terms from each of the functions $V^*_{i_1 i_2 \cdots i_t}$ and $V^*_{j_1 j_2 \cdots j_t}$, are extracted, that is

$$\sum_{r_1 \neq r_2 \cdots \neq r_t} V_{i_1}^{r_1} V_{i_2}^{r_2} \cdots V_{i_t}^{r_t} \quad \text{and} \quad \sum_{s_1 \neq s_2 \cdots \neq s_t} V_{j_1}^{s_1} V_{j_2}^{s_2} \cdots V_{j_t}^{s_t},$$

then using Propositions 3.2 and 3.3 establishes that the expectation of the product of the remainders can be at most $O(n^{t-1})$. Furthermore

$$E\left[\sum_{r_1 \neq r_2 \cdots \neq r_t} V_{i_1}^{r_1} V_{i_2}^{r_2} \cdots V_{i_t}^{r_t} \sum_{s_1 \neq s_2 \cdots \neq s_t} V_{j_1}^{s_1} V_{j_2}^{s_2} \cdots V_{j_t}^{s_t}\right]$$

equals

$$E\left[\sum_{r_1 \neq r_2 \cdots \neq r_t} \sum_p \left(V_{i_1}^{r_1} V_{j_1}^{r_1}\right) \left(V_{i_2}^{r_2} V_{j_2}^{r_2}\right) \cdots \left(V_{i_t}^{r_t} V_{j_t}^{r_t}\right)\right] = \sum_{r_1 \neq r_2 \cdots \neq r_t} \left(\sum_p W_{i_1,j_1} W_{i_2,j_2} \cdots W_{i_t,j_t}\right)$$

where $W_{i,j}$ is defined as in (3.5). This is a straight forward generalization of (3.11). The fact that the functions V_i^r are mean zero and independent has been used here. Noting that the number of terms in the summation $\sum_{r_1 \neq r_2 \neq \cdots \neq r_t}$ is $\frac{n!}{(n-t)!}$ establishes the result. ∎

The relevance of identifying this structure in the matrix $E[\mathbf{V}_t^* \mathbf{V}_t^{*T}]$ is that a matrix with elements $\sum_p W_{i_1,j_1} W_{i_2,j_2} \cdots W_{i_t,j_t}$ has a simple Moore-Penrose generalized inverse, it is

$$1/(t!)^2 \sum_p W^{i_1,j_1} W^{i_2,j_2} \cdots W^{i_t,j_t} \tag{3.21}$$

where $W^{i,j}$ are the elements of the inverse of $W_{i,j}$. The proof involves checking the four defining conditions of the Moore-Penrose inverse.

Since the matrix $E[\mathbf{V}_t^*\mathbf{V}_t^{*T}]$ is of dimensions $s^t \times s^t$ its inversion may well be totally infeasible, but the previous remark shows that it will be possible to approximate $E[\mathbf{V}_t^*\mathbf{V}_t^{*T}]^+$ by only inverting an $s \times s$ matrix, and performing a series of multiplications.

A brief lemma is needed to establish the key result in this section which concerns the order of magnitude in the difference between successive projective scores as measured by the variance of their difference.

Lemma 3.4 *Let A and B be any two square matrices of compatible dimensions. If the diagonal elements of A are of order $O(n^r)$ with all off diagonal elements of order at most $O(n^r)$, and the diagonal elements of B are $O(n^s)$ with the off diagonal elements again of order at most $O(n^s)$, then tr(AB) has order at most $O(n^{r+s})$.*

Proof

By definition $\mathrm{tr}(AB) = \sum_i \sum_j a_{ij} b_{ji}$. The largest order of magnitude in any of these terms is $O(n^{r+s})$, and allowing the possibility that terms of this order may cancel establishes that $\mathrm{tr}(AB)$ is *at most* $O(n^{r+s})$. ■

We are now finally in a position to establish the first important result concerning the Bhattacharyya basis functions. There are many choices for the basis to construct the projective scores, a basis with first element $\Pi_\infty U_0$ would obviously have the property of achieving the U_∞ score in one step. In general $\Pi_\infty U_0$ is not known and the Bhattacharyya basis has the desirable attribute of producing a steady reduction in the order of magnitude of the norm of the difference between successive projective scores:

Theorem 3.5 *Let $U_1, U_2, \cdots, U_t$ and the extended information matrix M_t be defined as before. Then*

$$I_{U_1} = \mathrm{Var}(U_1) = n(i_{\psi\psi} - i_{\psi\lambda} i_{\lambda\lambda}^{-1} i_{\lambda\psi}), \qquad tr\,[\mathrm{Var}(\mathbf{U}_1)] = O(n),$$

and in general for $B_t = U_t - U_{t-1}, \qquad t = 1, 2, 3, \cdots$

$$tr[\mathrm{Var}(B_t)] = tr[\mathrm{Var}(U_{t-1}) - \mathrm{Var}(U_t)] = tr\left[I_{U_{t-1}} - I_{U_t}\right] = O(n^{2-t}),$$

where $i_{\psi\psi}, i_{\lambda\lambda}, i_{\psi\lambda}$ and $i_{\lambda\psi}$ are the elements of the usual Fishers Information matrix based on a sample of size 1.

Proof

The first part of the theorem just states the formula for the marginal Fisher's information matrix for the parameter of interest. The second part is established using the partial orthogonalization of the Bhattacharyya basis functions. Clearly

$$\mathrm{Var}(U_{t-1}) - \mathrm{Var}(U_t) = E[U_0 \mathbf{V}_t^{*T}] E[\mathbf{V}_t^* \mathbf{V}_t^{*T}]^+ E[U_0 \mathbf{V}_t^{*T}]^T.$$

The matrix $E[U_0 \mathbf{V}_t^{*T}]$ has terms all of order $O(n)$, and from results established in Section 6 of this chapter the matrix $E[\mathbf{V}_t^* \mathbf{V}_t^{*T}]^+$ has terms $O(n^{-t})$. Using the fact that

$$\mathrm{tr}(E[U_0 \mathbf{V}_t^{*T}] E[\mathbf{V}_t^* \mathbf{V}_t^{*T}]^+ E[U_0 \mathbf{V}_t^{*T}]^T) = \mathrm{tr}(E[\mathbf{V}_t^* \mathbf{V}_t^{*T}]^+ E[U_0 \mathbf{V}_t^{*T}]^T E[U_0 \mathbf{V}_t^{*T}]),$$

and denoting the matrix $E[\mathbf{V}_t^* \mathbf{V}_t^{*T}]^+$ as A and the matrix $E[U_0 \mathbf{V}_t^{*T}]^T E[U_0 \mathbf{V}_t^{*T}]$ as B, then B has diagonal elements of order $O(n^2)$. Furthermore, A certainly has diagonal elements of order $O(n^{-t})$ so that Lemma 3.4 applies, and $\mathrm{tr}(AB) = O(n^{2-t})$ as required. ∎

In conclusion, this theorem gives a representation of U_t as

$$U_t = U_1 + \sum_{j=2}^{t} B_j,$$

where the B_j are uncorrelated, and the stochastic order of the correction terms is quickly decreasing in n.

3.5 The tensor structure in the orthogonalized Bhattacharyya basis

This section establishes that the first two Bhattacharyya basis functions $\mathbf{V}_1$ and $\mathbf{V}_2^*$ are in fact tensors. The value of this result is that contractions of tensors to scalars are invariants, so that if a contraction of tensors to a scalar is identified then the scalar must be invariant. In order to clarify this idea a minimum of tensor terminology will be introduced.

The summary is taken from McCullagh (1987).

The following conventions will be adopted: there is an implied summation over any index appearing once as a superscript and once as a subscript, so

$$a_i b^i = \sum_{i=1}^{n} a_i b^i$$

Define $\delta_i^r = 1$ if $i = r$, and 0 otherwise, then

$$a_i^r \delta_j^i = a_j^r$$

and

$$a_i^r \delta_j^i b_k^j = a_i^r b_k^i$$

There are two essentially different types of tensors, covariant and contravariant (and a third category that contains mixtures of the covariant and contravariant), the difference being on how the functions transform with respect to a change in basis of the coordinate system. More precisely, let $\omega = \omega^{i_1 i_2 \cdots i_d}$ be a d-dimensional array whose elements are functions of a variable x. Consider the transformation $y = g(x)$ from x to new variables $y = y^1, \cdots, y^p$ and let $a_i^r \equiv a_i^r(x) = \partial y^r / \partial x^i$ have full rank for all x. If $\bar{\omega}$, the value of ω for the transformed variables satisfies

$$\bar{\omega}^{r_1 r_2 \cdots r_d} = a_{i_1}^{r_1} a_{i_2}^{r_2} \cdots a_{i_d}^{r_d} \omega^{i_1 i_2 \cdots i_d}$$

then ω is said to be a contravariant tensor.

The transformation law for covariant tensors is

$$\bar{\omega}_{r_1 r_2 \cdots r_d} = b_{r_1}^{i_1} b_{r_2}^{i_2} \cdots b_{r_d}^{i_d} \omega_{i_1 i_2 \cdots i_d}$$

where $b_r^i = \partial x^i / \partial y^r$, the matrix inverse of a_i^r, satisfies $a_i^r b_r^j = \delta_i^j = a_r^j b_i^r$.

The function $g(.)$ is an element of a group, and within this thesis this group will be that of non-linear invertible transformations. Here, we are thinking of transformations of the *parameters*, not the data. When an array of functions is referred to as a tensor, this only has meaning within the context of a group of transformations. This is the first time that transformation invariance has been considered, but more attention to it will be given in Chapter 5, where certain curvature interpretations of the results presented earlier will be given.

Multiplication and addition of tensors is straight forward. The operation of contraction, that is summing over pairs of indices, say $\gamma_{kl}^i = \gamma_{klj}^{ij}$, is important because any scalar that is obtained through contraction of a tensor is invariant, that is a quantity unaffected by a transformation within the transformation group under consideration.

Now consider a transformation from λ to μ, so that by the chain rule

$$\frac{\partial}{\partial \mu^r} f(\mathbf{x}, \psi, \lambda(\mu)) = \frac{\partial \lambda^1}{\partial \mu^r} \frac{\partial}{\partial \lambda^1} f + \frac{\partial \lambda^2}{\partial \mu^r} \frac{\partial}{\partial \lambda^2} f + \cdots + \frac{\partial \lambda^s}{\partial \mu^r} \frac{\partial}{\partial \lambda^s} f$$

and hence

$$\tilde{V}_r = \frac{\frac{\partial}{\partial \mu^r} f(\mathbf{x}, \psi, \mu)}{f(\mathbf{x}, \psi, \mu)} = \lambda_r^i V_i.$$

where $\lambda_i^r = \partial\lambda^r/\partial\mu^i$. It follows that the score is a covariant tensor of order 1.

Now define $W_{i,j} = E[V_i V_j]$ and $W_{rs,t} = E[V_{rs}V_t]$ and consider the function

$$V_{ij}^* = V_{ij} - W_{ij,k}W^{k,l}V_l \tag{3.22}$$

which is simply the orthogonalized version of V_{ij} (see 3.20), where $W_{i,j}W^{j,k} = \delta_i^k = W^{j,k}W_{i,j}$ (ie $W^{i,j}$ is the matrix inverse of $W_{i,j}$). It can be shown that V_{ij}^* is a covariant tensor of order 2, McCullagh (1987, p.19, Exercise 1.5).

3.6 Permutation invariant matrices

The purpose of this section is to make precise some of the earlier remarks concerning the order of magnitudes of the elements of the Moore-Penrose inverses referred to in Sections 2 and 3 of this chapter. The aim is to establish that the Moore-Penrose inverse of $E[\mathbf{V}_t^*\mathbf{V}_t^{*T}]$ can be represented as

$$\frac{(n-t)!}{t!}\sum_{r_1\neq r_2\cdots\neq r_t}\left(\sum_p W^{i_1,j_1}W^{i_2,j_2}\dots W^{i_t,j_t}\right) + E_n, \tag{3.23}$$

where the remainder term E_n is asymptotically negligible compared to the main term. This is the result that was claimed in Proposition 4 and in the comment beneath (3.12). The previous claims implicitly relied on the continuity of the Moore-Penrose inverse; that is, the following statement is assumed true: if E_n is a sequence of square matrices going to zero then for A, a fixed square matrix

$$(A + E_n)^+ \to A^+.$$

This assertion is true if A is of full rank, but if A is rank deficient then there is no guarantee that the statement remains true. In particular, consider the following example taken from Campbell and Meyer (1979):

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \qquad E_n = \begin{pmatrix} 0 & 0 \\ 0 & \frac{1}{n} \end{pmatrix}$$

then

$$(A + E_n)^+ = \begin{pmatrix} 1 & 0 \\ 0 & n \end{pmatrix} \qquad A^+ = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

so that clearly the desired result is not true. Theorem 10.4.1 in Campbell and Meyer (1979) gives necessary and sufficient conditions for $(A+E_n)^+ \to A^+$. The condition is that for some fixed n_0, for all $n > n_0$ the rank of $A+E_n$ is equal to the rank of A.

The following theorem presents a sufficient condition for rank of $A+E_n$ to equal the rank of A for all $n > n_0$.

Theorem 3.6 *Let E_n be in the column space of A for all n, and $E_n \to 0$, then there exists an n_0 such that rank $A + E_n$ = rank A for all $n > n_0$.*

Proof

If E_n is in the column space of A for all n, then there exists a sequence of matrices F_n such that $E_n = AF_n$ and $F_n \to 0$, so that

$$(A + E_n) = (A + AF_n) = A(I + F_n),$$

where I is the identity matrix. Since $F_n \to 0$ there exists an n_0 such that for all $n > n_0$, $I + F_n$ is non-singular. Using (Rao, 1973, p.30,Exercise 1.2(b)), which states that if B is square and non-singular then rank(AB) = rank$A$, it follows that rank $A + E_n$ = rank A as required. ■

To prove the result about the Moore-Penrose inverse of $E[\mathbf{V}_t^* \mathbf{V}_t^{*T}]$, that is (3.23), some definitions need to be introduced.

A square matrix, denoted by W, of dimensions $s^t \times s^t$ can be indexed by the pairs $\{i_1 i_2 \cdots i_t, j_1 j_2 \cdots j_t\}$ where $1 \leq i_k \leq s$ and $1 \leq j_k \leq s$ for $k = 1, 2 \cdots t$. Such a matrix will be called *permutation invariant* if

$$W_{p_i, p_j} = W_{p'_i, p'_j} \quad \text{for all permutations } p'_i, p'_j \text{ of } p_i \text{ and } p_j$$

Here a permutation p'_i refers to a permuatation if the indices $\{i_1, i_2, \cdots, i_t\}$ and p'_j to a permutation of the indices $\{j_1, j_2, \cdots, j_t\}$. Note that permutation invariance is not a generalization of symmetry, since symmetry involves interchanging subscripts between rows and columns. There is no need for a permutation invariant matrix to be symmetric.

Lemma 3.5 *If the permutation invariant matrix B can be decomposed into the sum of two matrices $A + E$, and A is permutation invariant then so must E be.*

The proof is immediate: assume to the contrary that E is not a permutation invariant matrix, then B cannot be permutation invariant which is a contradiction.

■

The $s^t \times s^t$ matrix $E[\mathbf{V}_t {\mathbf{V}_t}^T]$ is clearly permutation invariant since for an arbitary element $E[V_{i_1 i_2 \cdots i_t} V_{j_1 j_2 \cdots j_t}]$, if $\{i_1, i_2, \cdots, i_t\}$ is permuted then $V_{i_1 i_2 \cdots i_t}$ does not change because differentiation is invariant with respect to the order in which the derivatives are taken. The same argument applies to the matrix $E[\mathbf{V}_t^* \mathbf{V}_t^{*T}]$, so that it too is permutation invariant.

Recall from Proposition 3.4 that the matrix with $\{i_1 i_2 \cdots i_t, j_1 j_2 \cdots j_t\}$ element given by

$$\sum_{r_1 \neq r_2 \cdots \neq r_t} \left(\sum_p W_{i_1, j_1} W_{i_2, j_2} \cdots W_{i_t, j_t} \right)$$

can be extracted from $E[\mathbf{V}_t^* \mathbf{V}_t^{*T}]$. By its very construction this matrix is permutation invariant, so that the remainder

$$E_n = E[\mathbf{V}_t^* \mathbf{V}_t^{*T}] - \sum_{r_1 \neq r_2 \cdots \neq r_t} \left(\sum_p W_{i_1,j_1} W_{i_2,j_2} \cdots W_{i_t,j_t} \right) \tag{3.24}$$

is, by Lemma 3.5, permutation invariant.

The next stage of the the argument is to show that E_n is in the column space of $\left(\sum_p W_{i_1,j_1} W_{i_2,j_2} \cdots W_{i_t,j_t}\right)$. That is, a matrix F_n must be found such that

$$E_n = \left(\sum_p W_{i_1,j_1} W_{i_2,j_2} \cdots W_{i_t,j_t} \right) F_n.$$

The summation $\sum_{r_1 \neq r_2 \cdots \neq r_t}$ has been dropped here since it only contributes a multiplicative constant. The identification of F_n is established with the help of the following lemma:

Lemma 3.6

$$\left(\sum_p W_{i_1,j_1} W_{i_2,j_2} \cdots W_{i_t,j_t} \right) \left(\sum_p W^{j_1,k_1} W^{j_2,k_2} \cdots W^{j_t,k_t} \right) = t! \left(\sum_p \delta_{i_1}^{k_1} \delta_{i_2}^{k_2} \cdots \delta_{i_t}^{k_t} \right),$$

where δ_i^k is the (i,k) element of an $s \times s$ identity matrix, and $W^{i,j}$ is the matrix inverse of $W_{i,j}$, which is by assumption non-singular.

Proof

Counting argument. ■

Now, for any $s^t \times s^t$ permutation invariant matrix, H say,

$$\frac{1}{t!} \left(\sum_p \delta_{i_1}^{k_1} \delta_{i_2}^{k_2} \cdots \delta_{i_t}^{k_t} \right) H = H$$

so that $\frac{1}{t!}\left(\sum_p \delta_{i_1}^{k_1} \delta_{i_2}^{k_2} \cdots \delta_{i_t}^{k_t}\right)$ acts just like an identity matrix on permutation invariant matrices.

Consequently

$$\left(\sum_p W_{i_1,j_1} W_{i_2,j_2} \cdots W_{i_t,j_t}\right) \left(\sum_p W^{j_1,k_1} W^{j_2,k_2} \ldots W^{j_t,k_t}\right) E_n = (t!)^2 E_n,$$

so that setting

$$F_n = \frac{1}{(t!)^2} \left(\sum_p W^{j_1,k_1} W^{j_2,k_2} \ldots W^{j_t,k_t}\right) E_n,$$

establishes that E_n is indeed in the column space of $\left(\sum_p W_{i_1,j_1} W_{i_2,j_2} \cdots W_{i_t,j_t}\right)$.

Taking out the factor of $\frac{n!}{(n-t)!}$ from $E[V_t^* V_t^{*T}]$ means that it can be written as

$$E[V_t^* V_t^{*T}] = \frac{n!}{(n-t)!} \left[\left(\sum_p W_{i_1,j_1} W_{i_2,j_2} \cdots W_{i_t,j_t}\right) + E_{n,i_1 i_2 \cdots i_t, j_1 j_2 \cdots j_t}\right] \tag{3.25}$$

where $E_n \to 0$ and E_n is in the column space of $\left(\sum_p W_{i_1,j_1} W_{i_2,j_2} \cdots W_{i_t,j_t}\right)$.

Putting all of these results together gives:

Theorem 3.7

$$\frac{(n-t)!}{n!} E[V_t^* V_t^{*T}]^+ = \left(\sum_p W_{i_1,j_1} W_{i_2,j_2} \cdots W_{i_t,j_t}\right) + o(1).$$

Proof

$$E[V_t^* V_t^{*T}] = \frac{n!}{(n-t)!} \left[\left(\sum_p W_{i_1,j_1} W_{i_2,j_2} \cdots W_{i_t,j_t}\right) + E_{n,i_1 i_2 \cdots i_t, j_1 j_2 \cdots j_t}\right]$$

where the remainder term, E_n, is defined in (3.24). Consequently

$$\frac{(n-t)!}{n!} E[V_t^* V_t^{*T}]^+ = \left[\left(\sum_p W_{i_1,j_1} W_{i_2,j_2} \cdots W_{i_t,j_t}\right) + E_{n,i_1 i_2 \cdots i_t, j_1 j_2 \cdots j_t}\right]^+.$$

But $E_{n,i_1 i_2 \cdots i_t, j_1 j_2 \cdots j_t}$ is in the column space of $\left(\sum_p W_{i_1,j_1} W_{i_2,j_2} \cdots W_{i_t,j_t}\right)$ which from (3.25) $\to 0$ as n gets large. Appealing now to Theorem 3.6 it follows that

$$\left[\left(\sum_p W_{i_1,j_1} W_{i_2,j_2} \cdots W_{i_t,j_t}\right) + E_{n,i_1 i_2 \cdots i_t, j_1 j_2 \cdots j_t}\right]^+ \to \sum_p W^{i_1,j_1} W^{i_2,j_2} \ldots W^{i_t,j_t}$$

which is the required result.