Conditional regression theorem

The following result shows the equivalence of projecting a variable into a single orthogonal basis (constructed sequentially), and using a biorthogonal combination of a simple and conditional basis. It also shows that – for first order terms – using biorthogonal bases yields the same result as partial regression analysis.

If we choose an (arbitrary) ordering of variables, and denote the polynomial in variable ϕ_j that is orthogonal to all variables numbered less than j as $\mathbf{P}^j_{<j}$, then we have:

Theorem (Conditional Regression)

$$\sum_{i=1}^{n} \left[\!\!\left[\begin{smallmatrix} f \\ \mathbf{P}^i_{<i} \end{smallmatrix} \right]\!\!\right] \mathbf{P}^i_{<i} = \sum_{i=1}^{n} \left[\!\!\left[\begin{smallmatrix} f \\ \mathbf{P}^i_{\bullet} \end{smallmatrix} \right]\!\!\right] \mathbf{P}^i \tag{1}$$

Proof

We can write $\mathbf{P}^n_{\bullet}$ as:

$$\mathbf{P}^n_{\bullet} = \mathbf{P}^n - \sum \left[\!\!\left[\begin{smallmatrix} \mathbf{P}^n \\ \mathbf{P}^i_{<i} \end{smallmatrix} \right]\!\!\right] \mathbf{P}^i_{<i} \tag{2}$$

The $\mathbf{P}^i_{<i}$ polynomials are all orthogonal to one another, though they are not orthogonal to $\mathbf{P}^n$. We can therefore write the variance of $\mathbf{P}^n_{\bullet}$ as:

$$[\![\mathbf{P}^n_{\bullet}]\!] = [\![\mathbf{P}^n]\!] + \sum_{i=1}^{n-1} \left[\!\!\left[\begin{smallmatrix} \mathbf{P}^n \\ \mathbf{P}^i_{<i} \end{smallmatrix} \right]\!\!\right]^2 [\![\mathbf{P}^i_{<i}]\!] - 2 \sum_{i=1}^{n-1} \left[\!\!\left[\begin{smallmatrix} \mathbf{P}^n \\ \mathbf{P}^i_{<i} \end{smallmatrix} \right]\!\!\right] [\![\mathbf{P}^n, \mathbf{P}^i_{<i}]\!] \tag{3}$$

Lemma: Variance of conditional bases

For n variables, the variance of the n^{th} conditional basis is:

$$[\![\mathbf{P}^n_{\bullet}]\!] = [\![\mathbf{P}^n]\!] - \sum_{i=1}^{n-1} \frac{[\![\mathbf{P}^n, \mathbf{P}^i_{<i}]\!]^2}{[\![\mathbf{P}^i_{<i}]\!]} \tag{4}$$

Where $\mathbf{P}^i_{<i}$ is the polynomial for the i^{th} variable independent of all lower ones.

To prove Equation 1, we first show that it holds for two variables, then use induction to prove the general case.

Two variables

We need to show that, for a variable f:

$$\left[\!\!\left[\begin{smallmatrix} f \\ \mathbf{P}_1^2 \end{smallmatrix} \right]\!\!\right] \mathbf{P}_1^2 + \left[\!\!\left[\begin{smallmatrix} f \\ \mathbf{P}^1 \end{smallmatrix} \right]\!\!\right] \mathbf{P}^1 = \left[\!\!\left[\begin{smallmatrix} f \\ \mathbf{P}_1^2 \end{smallmatrix} \right]\!\!\right] \mathbf{P}^2 + \left[\!\!\left[\begin{smallmatrix} f \\ \mathbf{P}_2^1 \end{smallmatrix} \right]\!\!\right] \mathbf{P}^1 \tag{5}$$

For two variables, Equation 4 tells us that the variance in $\mathbf{P}_1^2$ is given by:

$$[\![\mathbf{P}_1^2]\!] = [\![\mathbf{P}^2]\!] - \frac{[\![\phi_1, \phi_2]\!]^2}{[\![\mathbf{P}^1]\!]} = \frac{|C|}{[\![\mathbf{P}^1]\!]} \tag{6}$$

where $|C|$ is the determinant of the covariance matrix for $\mathbf{P}^1$ and $\mathbf{P}^2$.

Similarly,

$$[\![\mathbf{P}_2^1]\!] = \frac{|C|}{[\![\mathbf{P}^2]\!]} \tag{7}$$

For two variables, the variances are thus related by:

$$[\![\mathbf{P}_2^1]\!]\,[\![\mathbf{P}^2]\!] = [\![\mathbf{P}_1^2]\!]\,[\![\mathbf{P}^1]\!] \tag{8}$$

Expanding $\mathbf{P}_1^2$, the lefthand side of Equation 5 can be written:

$$\begin{aligned} \left[\!\!\left[\begin{smallmatrix} f \\ \mathbf{P}_1^2 \end{smallmatrix} \right]\!\!\right] \mathbf{P}_1^2 + \left[\!\!\left[\begin{smallmatrix} f \\ \mathbf{P}^1 \end{smallmatrix} \right]\!\!\right] \mathbf{P}^1 &= \left[\!\!\left[\begin{smallmatrix} f \\ \mathbf{P}_1^2 \end{smallmatrix} \right]\!\!\right] \mathbf{P}^2 - \left[\!\!\left[\begin{smallmatrix} f \\ \mathbf{P}_1^2 \end{smallmatrix} \right]\!\!\right] \left[\!\!\left[\begin{smallmatrix} \mathbf{P}^2 \\ \mathbf{P}^1 \end{smallmatrix} \right]\!\!\right] \mathbf{P}^1 + \left[\!\!\left[\begin{smallmatrix} f \\ \mathbf{P}^1 \end{smallmatrix} \right]\!\!\right] \mathbf{P}^1 \\ &= \left[\!\!\left[\begin{smallmatrix} f \\ \mathbf{P}_1^2 \end{smallmatrix} \right]\!\!\right] \mathbf{P}^2 + \left(\left[\!\!\left[\begin{smallmatrix} f \\ \mathbf{P}^1 \end{smallmatrix} \right]\!\!\right] - \left[\!\!\left[\begin{smallmatrix} f \\ \mathbf{P}_1^2 \end{smallmatrix} \right]\!\!\right] \left[\!\!\left[\begin{smallmatrix} \mathbf{P}^2 \\ \mathbf{P}^1 \end{smallmatrix} \right]\!\!\right] \right) \mathbf{P}^1 \end{aligned} \tag{9}$$

To confirm Equation 5, it suffices to show that:

$$\left[\!\!\left[\begin{smallmatrix} f \\ \mathbf{P}^1 \end{smallmatrix} \right]\!\!\right] - \left[\!\!\left[\begin{smallmatrix} f \\ \mathbf{P}_1^2 \end{smallmatrix} \right]\!\!\right] \left[\!\!\left[\begin{smallmatrix} \mathbf{P}^2 \\ \mathbf{P}^1 \end{smallmatrix} \right]\!\!\right] = \left[\!\!\left[\begin{smallmatrix} f \\ \mathbf{P}_2^1 \end{smallmatrix} \right]\!\!\right] \tag{10}$$

Rewriting the regressions as covariances divided by variances, the term in parentheses in Equation 9 becomes:

$$\frac{[\![f, \mathbf{P}^1]\!]}{[\![\mathbf{P}^1]\!]} - \frac{[\![f, \mathbf{P}_1^2]\!]\,[\![\mathbf{P}^1, \mathbf{P}^2]\!]}{[\![\mathbf{P}_1^2]\!]\,[\![\mathbf{P}^1]\!]} \tag{11}$$

Using Equation 8 to rewrite the denominator of the second term, this becomes:

$$\frac{[\![f, \mathbf{P}^1]\!]}{[\![\mathbf{P}^1]\!]} - \frac{[\![f, \mathbf{P}^2_1]\!]\,[\![\mathbf{P}^1, \mathbf{P}^2]\!]}{[\![\mathbf{P}^1_2]\!]\,[\![\mathbf{P}^2]\!]} \tag{12}$$

Again using 8 to rewrite the denominator of the first term, we get:

$$\frac{[\![f, \mathbf{P}^1]\!]\,[\![\mathbf{P}^2_1]\!] - [\![f, \mathbf{P}^2_1]\!]\,[\![\mathbf{P}^1, \mathbf{P}^2]\!]}{[\![\mathbf{P}^1_2]\!]\,[\![\mathbf{P}^2]\!]} \tag{13}$$

Expanding both $\mathbf{P}^2_1$ terms in the numerator yields:

$$\frac{[\![f, \mathbf{P}^1]\!]\,[\![\mathbf{P}^2]\!] - [\![f, \mathbf{P}^1]\!]\,\frac{[\![\mathbf{P}^1,\mathbf{P}^2]\!]^2}{[\![\mathbf{P}^1]\!]} - [\![f, \mathbf{P}^2]\!]\,[\![\mathbf{P}^1, \mathbf{P}^2]\!] + [\![f, \mathbf{P}^1]\!]\,\frac{[\![\mathbf{P}^1,\mathbf{P}^2]\!]^2}{[\![\mathbf{P}^1]\!]}}{[\![\mathbf{P}^1_2]\!]\,[\![\mathbf{P}^2]\!]} \tag{14}$$

The 2^{nd} and 4^{th} terms in the numerator of 14 cancel one another, and the $[\![\mathbf{P}^2]\!]$ in the denominator cancels from the first term in the numerator, and converts the third term in the numerator into a regression. We thus are left with:

$$\frac{[\![f, \mathbf{P}^1]\!] - \left[\!\!\left[\begin{smallmatrix}\mathbf{P}^1\\ \mathbf{P}^2\end{smallmatrix}\right]\!\!\right] [\![f, \mathbf{P}^2]\!]}{[\![\mathbf{P}^1_2]\!]} = \frac{[\![f, \mathbf{P}^1_2]\!]}{[\![\mathbf{P}^1_2]\!]} = \left[\!\!\left[\begin{smallmatrix}f\\ \mathbf{P}^1_2\end{smallmatrix}\right]\!\!\right] \tag{15}$$

Which confirms Equation 10 and thus also Equation 5, proving the Theorem (1) for the case of two variables. □

Any number of variables

Having proven the two variable case, we proceed by induction. If Equation 1 works for the first $(n-1)$ variables, then the projection of f into the space of n variables can be written (to first order):

$$\left[\!\!\left[\begin{smallmatrix}f\\ \mathbf{P}^n_\bullet\end{smallmatrix}\right]\!\!\right] \mathbf{P}^n_\bullet + \sum_{i=1}^{n-1} \left[\!\!\left[\begin{smallmatrix}f\\ \mathbf{P}^i_{\neq n}\end{smallmatrix}\right]\!\!\right] \mathbf{P}^i \tag{16}$$

Where $\mathbf{P}^i_{\neq n}$ is the polynomial in ϕ_i that is orthogonal to all other variables except ϕ_n, and $\mathbf{P}^n_\bullet$ is the polynomial in ϕ_n orthogonal to all other variables. Similarly, $\mathbf{P}^n_\bullet$ itself can be expanded as:

$$\mathbf{P}^n_\bullet = \mathbf{P}^n - \sum_{i=1}^{n-1} \left[\!\!\left[\begin{matrix} \mathbf{P}^n \\ \mathbf{P}^i_{\neq i} \end{matrix} \right]\!\!\right] \mathbf{P}^i \tag{17}$$

Substituting 17 into 16 and grouping terms in $\mathbf{P}^i$, the projection of f into the space of all n variables can be written:

$$\left[\!\!\left[\begin{matrix} f \\ \mathbf{P}^n_\bullet \end{matrix} \right]\!\!\right] \mathbf{P}^n + \sum_{i=1}^{n-1} \left(\left[\!\!\left[\begin{matrix} f \\ \mathbf{P}^i_{\neq n} \end{matrix} \right]\!\!\right] - \left[\!\!\left[\begin{matrix} f \\ \mathbf{P}^n_\bullet \end{matrix} \right]\!\!\right] \left[\!\!\left[\begin{matrix} \mathbf{P}^n \\ \mathbf{P}^i_{\neq n} \end{matrix} \right]\!\!\right] \right) \mathbf{P}^i \tag{18}$$

To prove that Equation 1 works for n variables given that it works for $n-1$, it suffices to show that:

$$\left[\!\!\left[\begin{matrix} f \\ \mathbf{P}^i_{\neq n} \end{matrix} \right]\!\!\right] - \left[\!\!\left[\begin{matrix} f \\ \mathbf{P}^n_\bullet \end{matrix} \right]\!\!\right] \left[\!\!\left[\begin{matrix} \mathbf{P}^n \\ \mathbf{P}^i_{\neq n} \end{matrix} \right]\!\!\right] = \left[\!\!\left[\begin{matrix} f \\ \mathbf{P}^i_\bullet \end{matrix} \right]\!\!\right] \quad \forall\ i < n \tag{19}$$

All terms on the lefthand side of Equation 19 involve operations in the space that is orthogonal to all variables other than i and n. We can thus define new polynomials that inhabit this space:

$$\begin{aligned} Q^1 &= \mathbf{P}^i_{\neq n} \\ Q^2 &= \mathbf{P}^n_{\neq i} \\ Q^1_2 &= \mathbf{P}^i_\bullet \\ Q^2_1 &= \mathbf{P}^n_\bullet \end{aligned} \tag{20}$$

Noting that $[\![Q^2, Q^1]\!] = [\![\mathbf{P}^n, Q^1]\!]$, Equation 19 can be rewritten as:

$$\left[\!\!\left[\begin{matrix} f \\ Q^1 \end{matrix} \right]\!\!\right] - \left[\!\!\left[\begin{matrix} f \\ Q^2_1 \end{matrix} \right]\!\!\right] \left[\!\!\left[\begin{matrix} Q^2 \\ Q^1 \end{matrix} \right]\!\!\right] = \left[\!\!\left[\begin{matrix} f \\ Q^1_2 \end{matrix} \right]\!\!\right] \tag{21}$$

Equation 21 is identical to Equation 10, so the same reasoning as used in Equations 11 through 15 confirms it.

Having proved the 2 dimensional case above, this suffices to prove Equation 1 for any number of variables greater than 2. □

Relation to multiple regression

The conditional regression theorem (1) shows that using biorthogonal bases yields – for first order terms – the same result as using standard multiple (partial) regression analysis. For higher order terms, the biorthogonal approach yields terms that show the effects of each order independent of lower orders. It thus does not lead to lower order terms changing as we add higher orders.

If we look at terms of order j, the biorthogonal analysis including j and higher orders will yield the same j^{th} order coefficients (but not the same coefficients for orders $< j$) as would a multiple regression analysis that went only up to order j.0.

Relation to the covariance matrix

The standard method of multiple regression involves using the covariance matrix of all variables (Stuart et al., 1999). Our discussion above of the conditional regression theorem did not involve this covariance matrix (except fleetingly), but it will be informative to see how it relates to our results.

For a variable a influenced by a set of phenotypic characters, define $\widetilde{a}$ as the mean value of a for a particular set of phenotypic values (meaning that $\widetilde{a}$ is the best prediction of a that we can make given only phenotypes). We will proceed as though there are n phenotypes, but note that this includes higher oder polynomials of multiple phenotypes. We can write:

$$\widetilde{a} = \overline{a} + \sum_{i=1}^{n} \left[\!\!\left[\begin{matrix} \widetilde{a} \\ \mathbf{P}^i_\bullet \end{matrix} \right]\!\!\right] \mathbf{P}^i \tag{22}$$

Given Equation 22, we can write the vector of covariances between a and the various phenotype polynomials as:

$$\underbrace{\begin{bmatrix} [\![\widetilde{a}, \mathbf{P}^1]\!] \\ [\![\widetilde{a}, \mathbf{P}^2]\!] \\ \vdots \\ [\![\widetilde{a}, \mathbf{P}^n]\!] \end{bmatrix}}_{S} = \underbrace{\begin{bmatrix} [\![\mathbf{P}^1]\!] & [\![\mathbf{P}^1, \mathbf{P}^2]\!] & \cdots & [\![\mathbf{P}^1, \mathbf{P}^n]\!] \\ [\![\mathbf{P}^2, \mathbf{P}^1]\!] & [\![\mathbf{P}^2]\!] & \cdots & [\![\mathbf{P}^2, \mathbf{P}^n]\!] \\ \vdots & \vdots & \ddots & \vdots \\ [\![\mathbf{P}^n, \mathbf{P}^1]\!] & [\![\mathbf{P}^n, \mathbf{P}^2]\!] & \cdots & [\![\mathbf{P}^n]\!] \end{bmatrix}}_{C} \underbrace{\begin{bmatrix} \left[\!\!\left[\begin{matrix} \widetilde{a} \\ \mathbf{P}^1_\bullet \end{matrix} \right]\!\!\right] \\ \left[\!\!\left[\begin{matrix} \widetilde{a} \\ \mathbf{P}^2_\bullet \end{matrix} \right]\!\!\right] \\ \vdots \\ \left[\!\!\left[\begin{matrix} \widetilde{a} \\ \mathbf{P}^n_\bullet \end{matrix} \right]\!\!\right] \end{bmatrix}}_{\beta} \tag{23}$$

If we define the vector of $[\![\widetilde{a}, \mathbf{P}^i]\!]$ terms as S, the phenotypic covariance matrix as C, and the vector of regressions of $\widetilde{a}$ on the conditional bases as β, then we can solve for β as:

$$\beta = C^{-1} S \tag{24}$$

This confirms that β is the same as a vector of partial regressions. The following results show further relationships between our results and the covariance matrix. We will use these results later when extending the conditional regression theorem to phenotypes that are sequences.

Theorem (variance-determinant)

Define:
C = Covariance matrix for all traits.
$C_{\neg i}$ = Covariance matrix that excludes row i and column i.

Then, using $||$ for the determinant, the variance of the conditional basis $\mathbf{P}^i_\bullet$ is given by:

$$[\![\mathbf{P}^i_\bullet]\!] = \frac{|C|}{|C_{\neg i}|} \tag{25}$$

(Note that $|C_{\neg i}|$ is the minor corresponding to element C_{ii}).

Corollary

$$|C| = \prod_{i=1}^{n} [\![\mathbf{P}^i_{<i}]\!] \tag{26}$$

Proof

Because we are concerned with determinants, we can always move the variable i to the last row and column (by swapping row i with the last row and column i with the last column).

Break up the full covariance matrix, C, so as to isolate the submatrix (minor) corresponding to the covariance matrix for all variables excluding variable n:

$$C = \begin{bmatrix} C_{\neg n} & B \\ & \\ D & [\![\mathbf{P}^n]\!] \end{bmatrix} \tag{27}$$

Where B is the column vector corresponding to the n^{th} column of C but excluding the last term ($[\![\mathbf{P}^n]\!]$), and D is the row vector corresponding to the n^{th} row of C but excluding the last term:

$$B = \begin{bmatrix} [\![\mathbf{P}^1, \mathbf{P}^n]\!] \\ [\![\mathbf{P}^2, \mathbf{P}^n]\!] \\ \vdots \\ [\![\mathbf{P}^{n-1}, \mathbf{P}^n]\!] \end{bmatrix} \qquad D = \begin{bmatrix} [\![\mathbf{P}^n, \mathbf{P}^1]\!] & [\![\mathbf{P}^n, \mathbf{P}^2]\!] & \cdots & [\![\mathbf{P}^n, \mathbf{P}^{n-1}]\!] \end{bmatrix} \tag{28}$$

By Schur's Determinant Identity (noting that $\llbracket \mathbf{P}^n \rrbracket$ is a scalar and therefore is its own determinant):

$$|C| = \llbracket \mathbf{P}^n \rrbracket \cdot |C_{\neg n} - B \llbracket \mathbf{P}^n \rrbracket^{-1} D| \tag{29}$$

We can rewrite the determinant on the righthand side of Equation 29 using the Matrix Determinant Lemma (Ding & Zhou, 2007):

$$|C_{\neg n} - B \llbracket \mathbf{P}^n \rrbracket^{-1} D| = |C_{\neg n}| \cdot (1 - \llbracket \mathbf{P}^n \rrbracket^{-1} \langle D, C_{\neg n}^{-1} B \rangle) \tag{30}$$

where $\langle\ \rangle$ denotes inner product. Substituting 30 into 29 yields:

$$\begin{aligned} |C| &= \llbracket \mathbf{P}^n \rrbracket \cdot |C_{\neg n}| \cdot (1 - \llbracket \mathbf{P}^n \rrbracket^{-1} \langle D, C_{\neg n}^{-1} B \rangle) \\ &= |C_{\neg n}| \cdot (\llbracket \mathbf{P}^n \rrbracket - \langle D, C_{\neg n}^{-1} B \rangle) \end{aligned} \tag{31}$$

The terms on the righthand side of Equation 31 can be interpreted by noting that, from Equation 24:

$$C_{\neg n}^{-1} B = \begin{bmatrix} \left\llbracket \begin{matrix} \mathbf{P}^n \\ \mathbf{P}^1_{\neg 1, n} \end{matrix} \right\rrbracket \\ \left\llbracket \begin{matrix} \mathbf{P}^n \\ \mathbf{P}^2_{\neg 2, n} \end{matrix} \right\rrbracket \\ \vdots \\ \left\llbracket \begin{matrix} \mathbf{P}^n \\ \mathbf{P}^{n-1}_{< n-1} \end{matrix} \right\rrbracket \end{bmatrix} \tag{32}$$

From the fact that, for a constant k, $k \llbracket a, b \rrbracket = \llbracket a, kb \rrbracket$, we can write:

$$\begin{aligned} \langle D, C_{\neg n}^{-1} B \rangle &= \sum_{i=1}^{n-1} \left\llbracket \begin{matrix} \mathbf{P}^n \\ \mathbf{P}^i_{\neq i, n} \end{matrix} \right\rrbracket \llbracket \mathbf{P}^n, \mathbf{P}^i \rrbracket \\ &= \left\llbracket \mathbf{P}^n, \sum_{i=1}^{n-1} \left\llbracket \begin{matrix} \mathbf{P}^n \\ \mathbf{P}^i_{\neq i, n} \end{matrix} \right\rrbracket \mathbf{P}^i \right\rrbracket \end{aligned} \tag{33}$$

By the conditional regression theorem, the summation inside the covariance in Equation 33 is the projection of $\mathbf{P}^n$ into the space of all $\mathbf{P}^{i \neq n}$.

$$\begin{aligned}
\langle D, C_{\neg n}^{-1} B\rangle &= \left[\!\!\left[\mathbf{P}^n, \sum_{i=1}^{n-1} \left[\!\!\left[\begin{matrix}\mathbf{P}^n \\ \mathbf{P}^i_{<i}\end{matrix}\right]\!\!\right] \mathbf{P}^i_{<i}\right]\!\!\right] \\
&= \sum_{i=1}^{n-1} \left[\!\!\left[\begin{matrix}\mathbf{P}^n \\ \mathbf{P}^i_{<i}\end{matrix}\right]\!\!\right] [\![\mathbf{P}^n, \mathbf{P}^i_{<i}]\!] \\
&= \sum_{i=1}^{n-1} \frac{[\![\mathbf{P}^n, \mathbf{P}^i_{<i}]\!]^2}{[\![\mathbf{P}^i_{<i}]\!]}
\end{aligned} \tag{34}$$

Substituting 34 into 31, we get:

$$|C| = |C_{\neg n}| \cdot \left([\![\mathbf{P}^n]\!] - \sum_{i=1}^{n-1} \frac{[\![\mathbf{P}^n, \mathbf{P}^i_{<i}]\!]^2}{[\![\mathbf{P}^i_{<i}]\!]}\right) \tag{35}$$

By Equation 4, the term in parentheses in Equation 35 is equal to $[\![\mathbf{P}^n_\bullet]\!]$, so we have:

$$|C| = |C_{\neg n}| \cdot [\![\mathbf{P}^n_\bullet]\!] \tag{36}$$

Since any variable can be moved to position n, this confirms Equation 25. $\square$

From Equation 36 it follows that if we ignore variable n altogether, then the determinant of the remaining matrix is $|C_{\neg n}| = |C_{\neg n, n-1}| \, [\![\mathbf{P}^{n-1}_{<n-1}]\!]$. We can therefore write Equation 36 as:

$$|C| = |C_{\neg n, n-1}| \cdot [\![\mathbf{P}^{n-1}_{<n-1}]\!] \cdot [\![\mathbf{P}^n_\bullet]\!] \tag{37}$$

Continuing this process, and noting that the determinant of the matrix that contains only $[\![\mathbf{P}^1]\!]$ is just $[\![\mathbf{P}^1]\!]$, we get the corollary.

References

Ding, J. & Zhou, A. 2007: Eigenvalues of rank-one updated matrices with some applications. Applied Mathematics Letters 20(12):1223–1226.

Stuart, A.; Ord, J. K. & Arnold, S. 1999: Kendall's Advanced Theory of Statistics, 6^{th} Ed., Vol. 2A. Arnold, London.