The Geometry of Logit and Probit

This short note is meant as a supplement to Chapters 2 and 3 of *Spatial Models of Parliamentary Voting* and the notation and reference to figures in the text below is to those two chapters.

To recap, the normal vector is denoted as $\mathbf{N_j}$ and its *reflection* as $-\mathbf{N_j}$. The normal vector is perpendicular to the cutting plane. The cutting plane in two dimensions is defined by the equation

$$N_{j1}(W_1 - Z_{j1}) + N_{j2}(W_2 - Z_{j2}) = 0 \qquad (1)$$

where N_{j1} and N_{j2} are the components of the normal vector and (Z_{j1}, Z_{j2}) is the midpoint of the roll call outcomes (see Figure 2.11). Any point, (W_1, W_2), that satisfies the equation above lies on the cutting plane. For example, if the normal vector is (3, -2) and the roll call midpoint is (1, 0), then this produces the equation:

$$3(w_1 - 1) - 2(w_2 - 0) = 3w_1 - 3 - 2w_2 = 3w_1 - 2w_2 - 3 = 0 \quad \text{or}$$

$$3w_1 - 2w_2 = 3$$

so that (0, -3/2), (1/3, -1), (2, 3/2), etc., all lie on the plane.

In three dimensions the cutting plane is defined by the equation

$$N_{j1}(W_1 - Z_{j1}) + N_{j2}(W_2 - Z_{j2}) + N_{j3}(W_3 - Z_{j3}) = 0 \qquad (2)$$

Where, as above, any point, (W_1, W_2, W_3), that satisfies the equation above lies on the cutting plane. For example, if the normal vector is (2, -2, 1) and the roll call midpoint is (1, 1, 1), this produces the equation:

$$2(w_1 - 1) - 2(w_2 - 1) + (w_3 - 1) = 2w_1 - 2w_2 + w_3 - 1 = 0 \quad \text{or}$$

$$2w_1 - 2w_2 + w_3 = 1$$

so that (0, 0, 1), (1, 0, -1), (0, 1, 3), etc., all lie on the plane.

For s dimensions the cutting plane is defined by the vector equation

$$\mathbf{N_j}'(\mathbf{W} - \mathbf{Z_j}) = \mathbf{0} \qquad \text{(book, 2.8)} \qquad (3)$$

where $\mathbf{N_j}$ is the s by 1 *normal vector* such that $\mathbf{N_j}'\mathbf{N_j} = 1$, $\mathbf{W}$ and $\mathbf{Z_j}$ are s by 1 vectors and $\mathbf{0}$ is an s by 1 vector of zeroes. The normal vector is constrained to be of unit length for roll call voting work but it is a *general* vector in other applications.

In general, if $\mathbf{W_A}$ and $\mathbf{W_B}$ are both points in the plane, $\mathbf{N_j}'\mathbf{W_A} = \mathbf{N_j}'\mathbf{W_B} = c_j$, where c_j is a scalar constant. Geometrically, every point in the plane projects onto the same point on the line defined by the normal vector, $\mathbf{N_j}$ and its reflection $-\mathbf{N_j}$ (see Figure 2.11). This projection point for the general case ($\mathbf{N_j}$ not necessarily normalized to one; that is, $\mathbf{N_j}'\mathbf{N_j} = 1$):

$$\mathbf{M_j} = c_j \frac{\mathbf{N_j}}{\sum_{k=1}^{s} N_{jk}^2} \qquad \text{(book, 2.9)} \qquad (4)$$

To see this consider the example of the 3-dimensional plane above:

$$\begin{bmatrix} 2 & -2 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ \mathbf{w}_3 \end{bmatrix} = 2\mathbf{w}_1 - 2\mathbf{w}_2 + \mathbf{w}_3 = 1 \qquad (5)$$

We need to find the point on the normal vector [2 -2 1] that satisfies equation (5); that is, the point where the plane passes through the normal vector itself. Let k be a scalar constant. The solution is:

$$\begin{bmatrix} 2 & -2 & 1 \end{bmatrix} \begin{bmatrix} 2\mathbf{k} \\ -2\mathbf{k} \\ \mathbf{k} \end{bmatrix} = 4\mathbf{k} + 4\mathbf{k} + \mathbf{k} = 1, \ \ \textbf{so that k=}\frac{1}{9}$$

And the point is $\begin{bmatrix} \frac{2}{9} & -\frac{2}{9} & \frac{1}{9} \end{bmatrix}$ which is given in equation (4).

Note that, by construction, $\mathbf{N_j}'\mathbf{M_j} = c_j$. In addition, in the roll call context, because the midpoint of the Yea and Nay policy points, $\mathbf{Z_j}$, is on the cutting plane, it also projects to the point $\mathbf{M_j}$. The cutting plane passes through the line formed by the normal vector and its reflection (see Figure 2.11) at the point $\mathbf{M_j}$.

In the case of a simple probit analysis, the cutting plane consists of all possible legislator ideal points such that the probability of the corresponding legislator voting Yea or voting Nay is exactly .5; namely:

P(legislator i votes Yea) = P(legislator i votes Nay) =

$$\Phi\left(\frac{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_s X_{is}}{\sigma}\right) = 1 - \Phi\left(\frac{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_s X_{is}}{\sigma}\right) =$$

$$\Phi(0) = .5$$

Where $\Phi(.)$ is the distribution function for the normal and X_{i1} , X_{i2} , …, X_{is} are legislator i's coordinates on the s dimensions. Because the β's and σ cannot be separately identified, the usual assumption is to set $\sigma = 1$. The equation above reduces to:

$$\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_s X_{is} = 0 \quad \text{or}$$

$$\beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_s X_{is} = \tilde{\boldsymbol{\beta}}'\mathbf{X_i} = -\beta_0 \qquad (6)$$

where $\mathbf{X_i}$ is the s-length vector of legislator coordinates:

$$\mathbf{X_i} = \begin{bmatrix} \mathbf{X_{i1}} \\ \mathbf{X_{i2}} \\ \mathbf{X_{i3}} \\ \vdots \\ \mathbf{X_{is}} \end{bmatrix}$$

and $\tilde{\boldsymbol{\beta}}$ is an s-length vector of the coefficients β_1 , β_2 , β_3 , …, β_s; that is:

$$\tilde{\boldsymbol{\beta}} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \boldsymbol{\beta}_3 \\ \vdots \\ \boldsymbol{\beta}_s \end{bmatrix}$$

Note that the expression $\tilde{\boldsymbol{\beta}}'\mathbf{X_i} = -\beta_0$ is exactly the same as $\mathbf{N_j}'\mathbf{W} = c_\mathbf{j}$, which was used above. Namely, set $\mathbf{N_j} = \tilde{\boldsymbol{\beta}}$ and every point in the plane projects onto the point:

$$\mathbf{M_j} = -\beta_0 \frac{\tilde{\boldsymbol{\beta}}}{\sum_{k=1}^{s} \tilde{\beta}_k^2}. \qquad \text{(book, 2.10)} \qquad (7)$$

In other words, in a regular probit context the coefficients on the independent variables form a normal vector to a plane that passes through the point $-\beta_0 \frac{\tilde{\boldsymbol{\beta}}}{\sum_{k=1}^{s} \tilde{\beta}_k^2}$. Note that this point is fixed with regard to σ. In particular,

$$\frac{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_s X_{is}}{\sigma} = \frac{\beta_0}{\sigma} + \frac{1}{\sigma}\tilde{\beta}' X_i$$

So that

$$-\beta_0 \frac{\tilde{\boldsymbol{\beta}}}{\sum_{k=1}^{s} \tilde{\beta}_k^2} = -\left(\frac{\beta_0}{\sigma}\right)\left(\frac{\frac{1}{\sigma}\tilde{\boldsymbol{\beta}}}{\frac{1}{\sigma^2}\sum_{k=1}^{s} \tilde{\beta}_k^2}\right) \qquad (8)$$

When there is *complete separation* – that is, no error, this plays an important role below.

The simple logit case is identical to probit when σ = 1. The logit probabilities are:

$$\frac{e^{(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + ... + \beta_s X_{is})}}{1 + e^{(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + ... + \beta_s X_{is})}} = \frac{1}{1 + e^{(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + ... + \beta_s X_{is})}} = .5$$

Canceling out the denominator and taking the natural log of both sides yields the same equation as probit:

$$\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_s X_{is} = 0$$

In both probit and logit the coefficients on the independent variables form a normal vector to a plane that passes through the point $-\beta_0 \frac{\tilde{\boldsymbol{\beta}}}{\sum_{k=1}^{s} \tilde{\beta}_k^2}$.

The cosine of the angle between the normal vectors from probit and logit should be very close to one. That is:

$$|\cos\theta| = \left| \frac{\tilde{\boldsymbol{\beta}}_{\mathbf{P}}' \tilde{\boldsymbol{\beta}}_{\mathbf{L}}}{\left\|\tilde{\boldsymbol{\beta}}_{\mathbf{P}}\right\| \left\|\tilde{\boldsymbol{\beta}}_{\mathbf{L}}\right\|} \right| \approx 1 \qquad (9)$$

where θ is the angle between the two normal vectors, and ||.|| is the corresponding norm of the normal vector. Computing |cosθ| is a useful check on the two estimation techniques. Several examples of this are shown in the Appendix.

The Geometry of Complete Separation (Perfect Voting)

To simplify the notation below let the normal probability density function be $\phi\left(\frac{\beta_0 + \tilde{\boldsymbol{\beta}}'\mathbf{X}_i}{\sigma}\right)$ and the distribution function be $\Phi\left(\frac{\beta_0 + \tilde{\boldsymbol{\beta}}'\mathbf{X}_i}{\sigma}\right)$ which I will simplify to just **φ** and **Φ**, respectively. I leave σ in the expressions because it is the source of the identification problem. Namely, as a practical matter, the cutting plane is identified (up to a slight wiggle, depending upon the number of points) but, because there is no error,

$\sigma \rightarrow 0$ and the *observed* coefficient vector explodes to entries of $+\infty$ or $-\infty$ because of the implicit division of the coefficients by σ.

In a standard presentation the i[th] row of the matrix of independent variables would be:

$$\begin{bmatrix} 1 \\ \mathbf{X_{i1}} \\ \mathbf{X_{i2}} \\ \mathbf{X_{i3}} \\ \vdots \\ \mathbf{X_{is}} \end{bmatrix}, \text{ and the coefficient vector would be } \begin{bmatrix} \boldsymbol{\beta}_0 \\ \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \boldsymbol{\beta}_3 \\ \vdots \\ \boldsymbol{\beta}_s \end{bmatrix}.$$

But for clarity of presentation I want to separate the intercept term, β_0, from the other s coefficients.

In the general context my use of $\mathbf{i \in Yea}$ and $\mathbf{i \in Nay}$ below corresponds to the binary dependent variable being "1" and "0", respectively. The first derivatives for the Probit Log-likelihood function are:

$$\frac{\partial \ln \mathbf{L}}{\partial \boldsymbol{\beta}} = \begin{bmatrix} \sum_{\mathbf{i \in Yea}} \frac{\phi}{\Phi} - \sum_{\mathbf{i \in Nay}} \frac{\phi}{1-\Phi} \\ \sum_{\mathbf{i \in Yea}} \frac{\phi}{\Phi} \mathbf{X_{i1}} - \sum_{\mathbf{i \in Nay}} \frac{\phi}{1-\Phi} \mathbf{X_{i1}} \\ \sum_{\mathbf{i \in Yea}} \frac{\phi}{\Phi} \mathbf{X_{i2}} - \sum_{\mathbf{i \in Nay}} \frac{\phi}{1-\Phi} \mathbf{X_{i2}} \\ \vdots \\ \sum_{\mathbf{i \in Yea}} \frac{\phi}{\Phi} \mathbf{X_{is}} - \sum_{\mathbf{i \in Nay}} \frac{\phi}{1-\Phi} \mathbf{X_{is}} \end{bmatrix} \tag{10}$$

Let a and b be indices from 1 to s. Treating β_0 separately, the second derivatives are

$$\frac{\partial^2 \ln L}{\partial \beta_0^2} = -\sum_{i \in Yea} \frac{\phi^2 + \phi(\Phi)\left(\frac{\beta_0 + \tilde{\beta}' X_i}{\sigma}\right)}{\Phi^2} - \sum_{i \in Nay} \frac{\phi^2 - \phi(1-\Phi)\left(\frac{\beta_0 + \tilde{\beta}' X_i}{\sigma}\right)}{(1-\Phi)^2} \tag{11A}$$

$$\frac{\partial^2 \ln L}{\partial \beta_0 \partial \beta_a} = -\sum_{i \in Yea} \frac{\phi^2 + \phi(\Phi)\left(\frac{\beta_0 + \tilde{\beta}' X_i}{\sigma}\right)}{\Phi^2} X_{ia} - \sum_{i \in Nay} \frac{\phi^2 - \phi(1-\Phi)\left(\frac{\beta_0 + \tilde{\beta}' X_i}{\sigma}\right)}{(1-\Phi)^2} X_{ia} \tag{11B}$$

And the remaining are:

$$\frac{\partial^2 \ln L}{\partial \beta_a \partial \beta_b} = -\sum_{i \in Yea} \frac{\phi^2 + \phi(\Phi)\left(\frac{\beta_0 + \tilde{\beta}' X_i}{\sigma}\right)}{\Phi^2} X_{ia} X_{ib} - \sum_{i \in Nay} \frac{\phi^2 - \phi(1-\Phi)\left(\frac{\beta_0 + \tilde{\beta}' X_i}{\sigma}\right)}{(1-\Phi)^2} X_{ia} X_{ib} \tag{11C}$$

Before turning the pure case of complete separation, consider the intermediate but vexing case where one of the independent variables is an indicator variable that *separates* with respect to the binary dependent variable. That is, whenever the indicator is "1" the corresponding value of the dependent variable is "1". Let the indicator variable be X_1. The first partial derivative is:

$$\frac{\partial \ln L}{\partial \beta_1} = \sum_{i \in Yea, x=1} \frac{\phi}{\Phi} X_{i1} + \sum_{i \in Yea, x=0} \frac{\phi}{\Phi} X_{i1} - \sum_{i \in Nay, x=0} \frac{\phi}{1-\Phi} X_{i1} = \sum_{i \in Yea, x=1} \frac{\phi}{\Phi} X_{i1} = 0 \tag{12}$$

Because multiplication by zero cancels the 2nd and 3rd terms and the case $i \in Nay, x = 1$, by definition, does not occur. Hence, for equation (12) to hold it must be the case that

because $X_{i1}=1$, $\sum_{i \in Yea, x=1} \frac{\phi\left(\frac{\beta_0 + \tilde{\beta}' X_i}{\sigma}\right)}{\Phi\left(\frac{\beta_0 + \tilde{\beta}' X_i}{\sigma}\right)} = \sum_{i \in Yea, x=1} \frac{\phi\left(\frac{\beta_1}{\sigma} + K_i\right)}{\Phi\left(\frac{\beta_1}{\sigma} + K_i\right)} \to 0$, where

$$K_i = \frac{\beta_0 + \beta_2 X_{i2} + \beta_3 X_{i3} + \ldots + \beta_s X_{is}}{\sigma}.$$

The $\mathbf{K_i}$ can be treated as constants so that for equation (12) to hold it must be the case that $\beta_1 \rightarrow +\infty$. Note that this means that the normal vector itself is not identified. With the $\mathbf{K_i}$ as constants so that σ is a constant, then there will be a different normal vector with every change in the value of β_1 so that the cutting plane is not identified.

The case of *complete separation* has a different geometry. By definition, there exists a plane that perfectly divides the Yeas ("1"s) from the Nays ("0"s). Let $\mathbf{N_j}$ be the s by 1 *normal vector* to this plane such that $\mathbf{N_j}'\mathbf{N_j} = 1$ (I do not need the "j" subscript here but I retain it for notational consistency). Hence, as discussed above, if $X_{\mathbf{A}}$ and $X_{\mathbf{B}}$ are both points in the plane, then $\mathbf{N_j}'\mathbf{X_A} = \mathbf{N_j}'\mathbf{X_B} = c_j$, where $c_{\mathbf{j}}$ is a scalar constant. For complete separation either we have:

$$\textbf{if } \mathbf{i} \in \textbf{Yea, then } \mathbf{N}_j'\mathbf{X}_i > \mathbf{c}_j\textbf{; and if } \mathbf{i} \in \textbf{Nay, then } \mathbf{N}_j'\mathbf{X}_i < \mathbf{c}_j \qquad (13)$$

or

$$\textbf{if } \mathbf{i} \in \textbf{Yea, then } \mathbf{N}_j'\mathbf{X}_i < \mathbf{c}_j\textbf{; and if } \mathbf{i} \in \textbf{Nay, then } \mathbf{N}_j'\mathbf{X}_i > \mathbf{c}_j$$

Without loss of generality I will assume that equation (13) is the correct polarity.

Using equation (6) above, this is equivalent to:

$$\textbf{if } \mathbf{i} \in \textbf{Yea, then } \frac{1}{\sigma}\tilde{\beta}'\mathbf{X}_i > -\frac{\beta_0}{\sigma}\textbf{; and if } \mathbf{i} \in \textbf{Nay, then } \frac{1}{\sigma}\tilde{\beta}'\mathbf{X}_i < -\frac{\beta_0}{\sigma} \qquad (14)$$

Or more simply:

$$\textbf{if } \mathbf{i} \in \textbf{Yea, then } \frac{\beta_0 + \beta_1\mathbf{X}_{i1} + \ldots + \beta_s\mathbf{X}_{is}}{\sigma} > 0\textbf{; and if } \mathbf{i} \in \textbf{Nay, then } \frac{\beta_0 + \beta_1\mathbf{X}_{i1} + \ldots + \beta_s\mathbf{X}_{is}}{\sigma} < 0$$

However, if this is true then the likelihood function forces the following:

$$\textbf{if } \mathbf{i} \in \textbf{Yea, then } \Phi\left(\frac{\beta_0 + \tilde{\beta}'\mathbf{X}_i}{\sigma}\right) \rightarrow 1\textbf{; and if } \mathbf{i} \in \textbf{Nay, then } \Phi\left(\frac{\beta_0 + \tilde{\beta}'\mathbf{X}_i}{\sigma}\right) \rightarrow 0$$

because $\sigma \rightarrow 0$.

Again, however, note that the cutting plane is identified up to a slight “wiggle” because the $\mathbf{X_i}$**’**s are fixed constants and, by definition, the plane perfectly divides the Yeas (“1”s) from the Nays (“0”s). This plane passes through the s-dimensional space of the independent variables and separates those cases corresponding to Yeas (“1”s) from those corresponding to Nays (“0”s).

In actual estimation, the likelihood function for the linear probit (and logit) problem is globally convex – the inverse of the Hessian (equation system 11 above) is negative definite. Hence, the gradient vector will quickly climb up the surface of the likelihood function to the global maximum. In the case of perfect separation what this means in practice is that the gradient vector “explodes” as it approaches the global maximum of zero; that is, the vector becomes infinitely long. Therefore, a straightforward solution to this problem is to apply the constraint $\tilde{\boldsymbol{\beta}}'\tilde{\boldsymbol{\beta}} = \mathbf{1}$ at each iteration and then adjusting the standard deviation term, σ, to compensate. When the standard deviation term begins to vanish, stop. It is then a simple matter to take the normal vector and find c_j with the Janice algorithm.

Note that this produces the coefficient vector corresponding to a *perfect* specification for this particular set of independent variables, **X,** and the binary dependent variable, **y.** However, since there is no error, there is no probability, and no standard errors.

Appendix

Here are two examples from the NES 2000 election study. The variables are whether or not the respondent voted (0=not voted, > 0 voted), and the independent variables are party (0 – 6), income (22 categories), race (0 = white; 1 = black), sex (0 = male, 1 = female), south (0=north, 1=south), education (1=high school, 2=some college, 3=college).

First Example: Simple Probit and Simple Logit

```
. probit voted party income race sex south education age

Iteration 0:   log likelihood =  -680.7266
Iteration 1:   log likelihood =  -607.9971
Iteration 2:   log likelihood =  -607.1895
Iteration 3:   log likelihood = -607.18902
Iteration 4:   log likelihood = -607.18902

Probit regression                                 Number of obs   =       1062
                                                  LR chi2(7)      =     147.08
                                                  Prob > chi2     =     0.0000
Log likelihood = -607.18902                       Pseudo R2       =     0.1080
------------------------------------------------------------------------------
       voted |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       party |    .019205   .0213884     0.90   0.369    -.0227155    .0611255
      income |   .0358093   .0132463     2.70   0.007     .0098471    .0617716
        race |   .0371024   .1733654     0.21   0.831    -.3026875    .3768924
         sex |  -3.23e-06   .0840525    -0.00   1.000    -.1647431    .1647366
       south |  -.1761923   .0892279    -1.97   0.048    -.3510758   -.0013089
   education |   .4239381   .0566663     7.48   0.000     .3128742    .5350021
         age |   .0194557   .0027814     6.99   0.000     .0140043    .0249071
       _cons |  -1.503988   .2022846    -7.44   0.000    -1.900458   -1.107517
------------------------------------------------------------------------------

. logit voted party income race sex south education age

Iteration 0:   log likelihood =  -680.7266
Iteration 1:   log likelihood = -607.95982
Iteration 2:   log likelihood =  -605.9353
Iteration 3:   log likelihood = -605.92532
Iteration 4:   log likelihood = -605.92532

Logistic regression                               Number of obs   =       1062
                                                  LR chi2(7)      =     149.60
                                                  Prob > chi2     =     0.0000
Log likelihood = -605.92532                       Pseudo R2       =     0.1099
------------------------------------------------------------------------------
       voted |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       party |   .0361497   .0358051     1.01   0.313     -.034027    .1063264
      income |   .0623748   .0229564     2.72   0.007     .0173811    .1073686
        race |    .076192    .285144     0.27   0.789    -.4826799    .6350639
         sex |   .0159128   .1406487     0.11   0.910    -.2597536    .2915792
       south |  -.3005962   .1483302    -2.03   0.043     -.591318   -.0098743
   education |    .720976   .0964645     7.47   0.000     .5319091    .9100428
         age |   .0335865   .0048184     6.97   0.000     .0241426    .0430303
       _cons |  -2.621774   .3514732    -7.46   0.000    -3.310649   -1.932899
```

	Probit Normalized (Normal Vector)	Logit Normalized (Normal Vector)	
Party	0.041498	0.045816	
Income	0.077377	0.078054	
Race	0.080171	0.096566	
Sex	-0.000007	0.020168	
South	-0.380719	-0.380975	
Education	0.916051	0.913764	
Age	0.042040	0.042568	

The correlation (Cosine) between the two normal vectors = 0.99976.

Second Example: Ordered Probit and Ordered Logit

```
. oprobit party voted income race sex south education age

Iteration 0:   log likelihood = -2055.2461
Iteration 1:   log likelihood = -1954.4476
Iteration 2:   log likelihood =  -1954.382
Iteration 3:   log likelihood =  -1954.382

Ordered probit regression                         Number of obs   =       1062
                                                  LR chi2(7)      =     201.73
                                                  Prob > chi2     =     0.0000
Log likelihood =  -1954.382                       Pseudo R2       =     0.0491

------------------------------------------------------------------------------
       party |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       voted |   .4241235   .0395748    10.72   0.000     .3465583    .5016887
      income |   .0176777   .0095894     1.84   0.065    -.0011172    .0364726
        race |  -.8669836   .1432798    -6.05   0.000    -1.147807   -.5861603
         sex |  -.0837943    .064829    -1.29   0.196    -.2108568    .0432682
       south |    .246657   .0693117     3.56   0.000     .1108085    .3825055
   education |  -.0019693    .043899    -0.04   0.964    -.0880097    .0840712
         age |  -.0066873   .0021422    -3.12   0.002     -.010886   -.0024886
-------------+----------------------------------------------------------------
       /cut1 |  -.7521825   .1517723                     -1.049651   -.4547142
       /cut2 |  -.2260099    .148625                     -.5173096    .0652898
       /cut3 |   .2007637   .1477327                     -.0887871    .4903144
       /cut4 |   .5218228   .1485304                      .2307086     .812937
       /cut5 |   .9618794   .1510958                      .6657371    1.258022
       /cut6 |   1.470619   .1557338                      1.165386    1.775852
------------------------------------------------------------------------------

. ologit party voted income race sex south education age

Iteration 0:   log likelihood = -2055.2461
Iteration 1:   log likelihood = -1958.0084
Iteration 2:   log likelihood = -1957.1315
Iteration 3:   log likelihood = -1957.1306

Ordered logistic regression                       Number of obs   =       1062
                                                  LR chi2(7)      =     196.23
                                                  Prob > chi2     =     0.0000
Log likelihood = -1957.1306                       Pseudo R2       =     0.0477

------------------------------------------------------------------------------
       party |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       voted |    .695314   .0682924    10.18   0.000     .5614633    .8291647
      income |   .0295464   .0160355     1.84   0.065    -.0018826    .0609755
        race |   -1.45739   .2404073    -6.06   0.000     -1.92858   -.9862005
         sex |  -.1442224   .1097141    -1.31   0.189     -.359258    .0708132
       south |   .3940648   .1171609     3.36   0.001     .1644337    .6236959
   education |  -.0127451   .0749615    -0.17   0.865     -.159667    .1341767
         age |  -.0115873   .0036647    -3.16   0.002    -.0187699   -.0044047
-------------+----------------------------------------------------------------
       /cut1 |   -1.31551   .2545206                     -1.814361   -.8166589
       /cut2 |  -.4210721   .2477715                     -.9066953    .0645512
       /cut3 |   .2779521   .2460457                     -.2042886    .7601929
       /cut4 |   .8034034   .2477555                      .3178115    1.288995
       /cut5 |   1.534033   .2529847                      1.038192    2.029874
       /cut6 |    2.41272   .2622729                      1.898674    2.926765
------------------------------------------------------------------------------
```

	Probit Normalized (Normal Vector)	Logit Normalized (Normal Vector)	
voted	0.424174	0.416670	
Income	0.017680	0.017706	
Race	-0.867086	-0.873346	
Sex	-0.083804	-0.086426	
South	0.246686	0.236145	
Education	-0.001970	-0.007638	
Age	-0.006688	-0.006944	

The correlation (Cosine) between the two normal vectors = 0.99995.