NOTES ON THE NOMINATE MODEL

(December, 2006)

Legislator i's (i=1,…,p) utility for the Yea outcome on roll call j (j=1,…,q) is:

$$U_{ijy} = u_{ijy} + \varepsilon_{ijy} = \boldsymbol{\beta} e^{\left(-\frac{1}{2}\sum_{k=1}^{s} d_{ijky}^{2}\right)} + \varepsilon_{ijy} \qquad (1)$$

where u_{ijy} is the deterministic portion of the utility function, ε_{ijy} is the stochastic portion, and d_{ijky}^{2} is the distance of legislator i to the Yea outcome on the k^{th} (k=1,…,s) dimension for roll call j:

$$d_{ijky}^{2} = (X_{ik} - O_{jky})^2 \qquad (2)$$

and X_{ik} is the i^{th} legislator's ideal point on dimension k, and O_{jky} is the "Yea" outcome location for the j^{th} roll call on the k^{th} dimension.

Note that u_{ijy} bears a strong family resemblance to a multivariate normal distribution with variance-covariance matrix, $\boldsymbol{\Sigma}$, equal to the s by s identity matrix, $\mathbf{I_s}$; namely:

$$u_{ijy} \sim \beta^{*} \frac{1}{(2\pi)^{\frac{s}{2}}} e^{\left(-\frac{1}{2}\sum_{k=1}^{s}\left(O_{jky} - X_{ik}\right)^{2}\right)} \qquad (3)$$

where β is simply a proportionality constant that functions as a "signal-to-noise" ratio; that is:

$$\beta = \beta^{*} \frac{1}{\left(2\pi\right)^{\frac{s}{2}}}$$

Although u_{ijy} is not a probability distribution it is instructive to look at what is implied about the underlying metric by equation (3). Ignoring β*, if $\mathbf{\Sigma} = \mathbf{I_s}$ and the vector of means, **μ**, is the s length ideal point vector, $\mathbf{X_i}$, then by definition for large samples:

$$\frac{1}{q}\sum_{j=1}^{q}\left(O_{ky} - X_{ik}J_q\right)'\left(O_{ky} - X_{ik}J_q\right) = 1$$

$$\frac{1}{q}\sum_{j=1}^{q}\left(O_{ky} - X_{ik}J_q\right)'\left(O_{\ell y} - X_{i\ell}J_q\right) = 0,\ \ k \neq \ell$$

where $\mathbf{O_{ky}}$ is the q-length vector of Yea policy outcomes on dimension k, J_q is q-length vector of "1"s, and X_{ik} and $X_{i\ell}$ are the i[th] legislator's ideal point on dimensions k and ℓ respectively.

In other words, treating the utility function as if it were a multivariate normal probability density function means that the outcomes are concentrated within a hypersphere of radius 2 centered on the ideal point *for reasonable values of s*. For example, in two dimensions using polar coordinates, 86.5 percent of the outcomes will be within a circle of radius 2:

$$\frac{1}{2\pi}\int_0^2\int_0^{2\pi} re^{-\frac{r^2}{2}}d\theta dr = \int_0^2 re^{-\frac{r^2}{2}}dr = -e^{-\frac{r^2}{2}}\Big|_0^2 = -\frac{1}{e^2} - -1 = 1 - 0.135 = 0.865$$

Note that in one dimension this is 95.45 percent. As the number of dimensions increases the distribution becomes "squished" down so that the radius of the hypersphere containing a fixed percentage of the outcomes, for example, 95 percent, must *increase.* (Unfortunately, the multivariate normal can only be integrated in two dimensions.) As a practical matter this is not much of a problem because most applications are in one and two dimensions. However, for higher dimensional scalings this characteristic must be taken into account.

In any event it is not sensible to assume that the policy outcome points are distributed in some symmetric form around a *particular* legislator. The legislator ideal points themselves are dispersed over the underlying evaluative dimensions so that extreme legislators cannot be viewed as *means* of multivariate normal distributions of policy outcomes. In addition, the dispersion of the outcomes must be greater than the dispersion of the legislators because, for lopsided roll calls, the winning alternative is acceptable to a large majority of legislators so that the losing alternative must be relatively extreme. Specifically,

$$\frac{1}{q}\sum_{j=1}^{q}\left(O_{ky}-\bar{O}_{ky}J_q\right)'\left(O_{ky}-\bar{O}_{ky}J_q\right) \approx \frac{1}{q}\sum_{j=1}^{q}\left(O_{kn}-\bar{O}_{kn}J_q\right)'\left(O_{kn}-\bar{O}_{kn}J_q\right) >$$

$$\frac{1}{p}\sum_{j=1}^{p}\left(X_k-\bar{X}_kJ_p\right)'\left(X_k-\bar{X}_kJ_p\right) \ \forall \text{ k=1,...,s}$$

Where $\bar{O}_{ky}$ is the mean of the q Yea policy outcomes on dimension k, $\bar{X}_k$ is the mean of the p legislator ideal points on dimension k, and J_p is a p length vector of "1"s.

However, the dispersion of the *winning* alternatives should be approximately the same as the dispersion of the legislators. Let $\mathbf{O_{kw}}$ be the q-length vector of winning policy outcomes on dimension k. Then

$$\frac{1}{q}\sum_{j=1}^{q}\left(O_{kw}-\bar{O}_{kw}J_q\right)'\left(O_{kw}-\bar{O}_{kw}J_q\right) \approx \frac{1}{p}\sum_{j=1}^{p}\left(X_k-\bar{X}_kJ_p\right)'\left(X_k-\bar{X}_kJ_p\right) \ \forall \text{ k=1,...,s} \quad (4)$$

Ignoring the problem of constraints on the distribution of the ideal points and/or the policy outcomes for the time being, note that because there is no *absolute* metric the same level of utility, U_{ijy} , can be produced be either fixing the scale of the deterministic utility, u_{ijy} , and varying the standard deviation of the error, ε_{ijy} , that is:

$$\varepsilon_{ijy} \sim N(0, \frac{\sigma^2}{2}) \quad \text{so that} \quad \varepsilon_{ijn} - \varepsilon_{ijy} \sim N(0, \sigma^2)$$

Alternatively, σ^2 can be fixed and the relative weight of u_{ijy} in the overall utility, U_{ijy} , can be adjusted by increasing/decreasing β. In other words, without loss of generality we can assume that:

$$\varepsilon_{ijn} - \varepsilon_{ijy} \sim N(0, 1) \tag{5}$$

This implies that the distribution of the difference between the latent utilities for Yea and Nay is normal with mean $u_{ijy} - u_{ijn}$ and variance 1; that is

$$y^*_{ij} = U_{ijy} - U_{ijn} = u_{ijy} - u_{ijn} + \varepsilon_{ijn} - \varepsilon_{ijy} = \beta\left\{e^{\left(-\frac{1}{2}\sum_{k=1}^{s} d^2_{ijky}\right)} - e^{\left(-\frac{1}{2}\sum_{k=1}^{s} d^2_{ijkn}\right)}\right\} + \varepsilon_{ijn} - \varepsilon_{ijy}$$

$$\sim N(u_{ijy} - u_{ijn}, 1) \tag{6}$$

where y^*_{ij} is the difference between the latent utilities.

The probability that legislator i votes Yea on the j^{th} roll call is:

$P_{ijy} = P(U_{ijy} > U_{ijn}) = P(\varepsilon_{ijn} - \varepsilon_{ijy} < u_{ijy} - u_{ijn}) = \Phi(u_{ijy} - u_{ijn}) =$

$$\Phi\left[\beta\left\{e^{\left(-\frac{1}{2}\sum_{k=1}^{s} d^2_{ijky}\right)} - e^{\left(-\frac{1}{2}\sum_{k=1}^{s} d^2_{ijkn}\right)}\right\}\right] \tag{7}$$

Let **Y** be the p by q matrix of observed Yea/Nay choices and let $\mathbf{Y}^*$ be the p by q matrix of unobserved latent utility differences. From a classical perspective the *joint probability distribution of the sample* is:

$$\mathbf{f}(\mathbf{Y}^* \mid u_{ijy}\text{-}u_{ijn}) = \prod_{i=1}^{p}\prod_{j=1}^{q}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(y^*_{ij}-(u_{ijy}-u_{ijn}))^2} \tag{8}$$

Where $\mathbf{Y}^*$ is the p by q matrix of latent utility differences. Note that equation (8) is not a typical joint p.d. of the sample. Technically, a random sample is a set of independent and identically distributed random variables so that the joint p.d. of the sample is (DeGroot, 1986, p. 316):

$$f_n(\mathbf{X_1}, \mathbf{X_2}, \ldots, \mathbf{X_n} \mid \boldsymbol{\theta}) = f(\mathbf{X_1} \mid \boldsymbol{\theta})\, f(\mathbf{X_2} \mid \boldsymbol{\theta}) \ldots f(\mathbf{X_n} \mid \boldsymbol{\theta})$$

where $f(\mathbf{X} \mid \boldsymbol{\theta})$ is the distribution from which the random sample, $\mathbf{X_1}, \mathbf{X_2}, \ldots, \mathbf{X_n}$, is drawn. In contrast each of the pq elements of $\mathbf{Y}^*$ is a random sample of size one from the corresponding $N(u_{ijy} - u_{ijn}, 1)$ distribution. The joint p.d. is a pq-variate normal distribution with variance-covariance matrix $\mathbf{I_{pq}}$:

$$\mathbf{f}(\mathbf{Y}^* \mid u_{ijy}\text{-}u_{ijn}) =$$

$$\frac{1}{(2\pi)^{\frac{pq}{2}}} e^{-\frac{1}{2}\{(y_{11}^*-(u_{11y}-u_{11n}))^2+(y_{12}^*-(u_{12y}-u_{12n}))^2+\ldots+(y_{pq}^*-(u_{pqy}-u_{pqn}))^2\}} \qquad \textbf{(8A)}$$

To see that equation (8) is indeed a legal probability distribution note that:

$$\mathbf{f}(\mathbf{Y}^* \mid u_{ijy}\text{-}u_{ijn}) \geq 0 \text{ for all } y_{ij}^*,\ -\infty < y_{ij}^* < +\infty$$

and

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\ldots\int_{-\infty}^{+\infty} f(\mathrm{Y}^* \mid u_{ijy} - u_{ijn})\, dy_{11}^* dy_{21}^* \ldots dy_{pq}^* =$$

$$\left\{\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y_{11}^*-(u_{11y}-u_{11n}))^2} dy_{11}^*\right\}\left\{\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y_{21}^*-(u_{21y}-u_{21n}))^2} dy_{21}^*\right\}\ldots$$
$$\left\{\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y_{pq}^*-(u_{pqy}-u_{pqn}))^2} dy_{pq}^*\right\} = 1*1*\ldots*1 = 1$$

In the joint p.d. of the sample, $\mathbf{f}(\mathbf{Y}^* \mid u_{ijy}\text{-}u_{ijn})$**,** the y_{ij}^* are the random variables and the ps+2qs+1 parameters -- X_{i1} , X_{i2} , …, X_{is} , the qs Yea outcome coordinates – O_{j1y} , O_{j2y} ,…, O_{jsy} , the qs Nay outcome coordinates -- O_{j1n} , O_{j2n} ,…, O_{jsn} , and β – are fixed constants. Following DeGroot (1986, p. 317), if we regard $\mathbf{f}(\mathbf{Y}^* \mid u_{ijy}\text{-}u_{ijn})$ as a function of the parameters for given values of the y^*_{ij} then it is a *likelihood function*; that is

$$\mathbf{L}^*(u_{ijy}\text{-}u_{ijn} \mid \mathbf{Y}^*) =$$

$$\frac{1}{(2\pi)^{\frac{pq}{2}}} e^{-\frac{1}{2}\{(y^*_{11}-(u_{11y}-u_{11n}))^2+(y^*_{12}-(u_{12y}-u_{12n}))^2+...+(y^*_{pq}-(u_{pqy}-u_{pqn}))^2\}} \qquad \textbf{(9)}$$

Which is identical to equation (8) only now the pq y^*_{ij} are *observed* and the ps+2qs+1 parameters are *variables* (but *not* random variables), and the problem is to find values of the parameters that maximize equation (9).

Equation (8) is a probability distribution over the pq dimensional hyperplane with dimensions corresponding to the y^*_{ij}. Equation (9) is a function defined over the ps+2qs+1 dimensional hyperplane with dimensions corresponding to the ps legislator coordinates -- X_{i1} , X_{i2} , …, X_{is} , the qs Yea outcome coordinates – O_{j1y} , O_{j2y} ,…, O_{jsy} , the qs Nay outcome coordinates -- O_{j1n} , O_{j2n} ,…, O_{jsn} , and β. Although equation (9) is not a probability distribution it is the case that it is above zero over the ps+2qs+1 hyperplane; that is:

$$\mathbf{L}^*(u_{ijy}\text{-}u_{ijn} \mid \mathbf{Y}^*) \geq 0, \;\; 0 < \beta < +\infty, \;\; -\infty < X_{ik} , O_{jky} , O_{jkn} < +\infty$$

In addition, the hypervolume underneath the function is almost certainly finite:

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}...\int_{0}^{+\infty} L^*(\beta\left\{e^{\left(-\frac{1}{2}\sum_{k=1}^{s}d^2_{ijky}\right)} - e^{\left(-\frac{1}{2}\sum_{k=1}^{s}d^2_{ijkn}\right)}\right\}|\mathrm{Y}^*)d\beta dX_{11}...dX_{pk}dO_{11y}...dO_{qky}...dO_{11n}...dO_{qkn} =$$

$$K^* < +\infty \qquad (10)$$

Because as $\beta \rightarrow +\infty$ clearly $\mathbf{L}^*(u_{ijy}\text{-}u_{ijn} \mid \mathbf{Y}^*) \rightarrow 0$; and as the absolute value of any of the legislator and roll call parameters becomes large the likelihood function goes to zero; that is, as $|X_{ik}| \rightarrow +\infty$ clearly $\mathbf{L}^*(u_{ijy}-u_{ijn} \mid \mathbf{Y}^*) \rightarrow 0$. **L*** is shaped like a multivariate normal in that it is quasi-concave along each dimension and asymptotes towards zero fairly quickly. However, I have no formal proof that the hypervolume is finite.

The fact that $\mathbf{L}^*(u_{ijy}-u_{ijn} \mid \mathbf{Y}^*)$ is everywhere non-negative and almost certainly has a finite integral is important because the same simulation methods used to find the values of parameters for Bayesian posterior distributions – Metropolis-Hastings sampling and Gibbs sampling – can be utilized with **L*** because, for all intents and purposes, it can be treated as a probability distribution. This property will figure in the discussion of the actual likelihood function, $\mathbf{L}(u_{ijy}-u_{ijn} \mid \mathbf{Y})$, below.

Unfortunately, the latent utility differences are not observed and we do not have any simple expression for the joint probability distribution for the sample of discrete choices -- $f(\mathbf{Y} \mid u_{ijy}-u_{ijn})$. However, it is easy to write down the distribution corresponding to any particular choice, that is, $f_{ij}(y_{ij} \mid u_{ijy}-u_{ijn})$. The product of these pq distributions is proportion to the joint p.d. of the sample and the likelihood function. Specifically, let:

$$y_{ij} = \begin{cases} 1 \text{ (Yea) if } y^*_{ij} > 0 \\ 0 \text{ (Nay) if } y^*_{ij} \leq 0 \end{cases} \text{ so that } \begin{cases} P(y^*_{ij} > 0) = \boldsymbol{\Phi}(u_{ijy} - u_{ijn}) \\ P(y^*_{ij} \leq 0) = 1 - \boldsymbol{\Phi}(u_{ijy} - u_{ijn}) \end{cases}$$

If the y_{ij} are independent Bernoulli random variables, that is:

$$f_{ij}(y_{ij}|u_{ijy} - u_{ijn}) \sim \text{Bernoulli}(\boldsymbol{\Phi}(u_{ijy} - u_{ijn}))$$

then

$$f(\mathbf{Y} \mid u_{ijy}-u_{ijn}) \propto \prod_{i=1}^{p}\prod_{j=1}^{q} f_{ij}(y_{ij} \mid u_{ijy} - u_{ijn}) = L(u_{ijy}-u_{ijn} \mid \mathbf{Y}) =$$

$$\prod_{i=1}^{p}\prod_{j=1}^{q}[\Phi(u_{ijy} - u_{ijn})]^{y_{ij}}[1-\Phi(u_{ijy} - u_{ijn})]^{(1-y_{ij})} = \prod_{i=1}^{p}\prod_{j=1}^{q}\prod_{\tau=1}^{2} P_{ij\tau}^{C_{ij\tau}} \qquad (11)$$

where τ is the index for Yea and Nay, $P_{ij\tau}$ is the probability of voting for choice τ, and $C_{ij\tau}$= 1 if the legislator's actual choice is τ and zero otherwise. (This representation is convenient for working with the derivatives and for multi-choice situations. More on this below.)

Note that f(**Y |** u_{ijy}-u_{ijn}**)** is a *discrete* distribution with 2^{pq} possible outcomes. By definition, the ps+2qs+1 parameters are fixed constants and the $\mathbf{y_{ij}}$ are the random variables. Hence, we can apply standard probability theory to find the proportionality constant:

$$\sum_{y_{11}=0}^{1}\sum_{y_{12}=0}^{1}\cdots\sum_{y_{pq}=0}^{1}\left\{\prod_{i=1}^{p}\prod_{j=1}^{q} f_{ij}(y_{ij} \mid u_{ijy} - u_{ijn})\right\} = K$$

So that

$$f(Y \mid u_{ijy} - u_{ijn}) = \frac{1}{K}\left\{\prod_{i=1}^{p}\prod_{j=1}^{q} f_{ij}(y_{ij} \mid u_{ijy} - u_{ijn})\right\}$$

Fortunately, knowing the value of K is not important and does not affect the analysis of the likelihood function, L(u_{ijy}-u_{ijn} **| Y).** Technically, the likelihood function has exactly the same expression as f(**Y** | u_{ijy}-u_{ijn}). In this case the 1/K is missing. This has no effect because it is as if the true likelihood function were multiplied by K. When gradient methods are applied to the likelihood function all that matters is the relative heights of the function. In addition, when logs are taken the proportionality constant becomes an additive constant and plays no role in the estimation.

The joint p.d. of the sample, f(**Y |** u_{ijy}-u_{ijn}**),** is a *discrete* probability distribution with 2^{pq} possible outcomes. The likelihood function, L(u_{ijy}-u_{ijn} **| Y**), is a continuous distribution over the ps+2qs+1 dimensional hyperplane with dimensions corresponding to the ps legislator coordinates, the 2qs outcome coordinates, and β. Although equation (11) is not a probability distribution it is the case that it is above zero over the ps+2qs+1 hyperplane; that is:

$$\mathbf{L}(\, u_{ijy}\text{-}u_{ijn} \mid \mathbf{Y}\,) \geq 0,\;\; 0 < \beta < +\infty,\;\; -\infty < X_{ik}\,,\, O_{jky}\,,\, O_{jkn} < +\infty$$

Unfortunately the hypervolume underneath the function is *not* finite; that is:

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{0}^{+\infty} L\left(\beta\left\{e^{\left(-\frac{1}{2}\sum_{k=1}^{s} d_{ijky}^{2}\right)} - e^{\left(-\frac{1}{2}\sum_{k=1}^{s} d_{ijkn}^{2}\right)}\right\}\middle|\mathrm{Y}\right) d\beta dX_{11}\ldots dX_{pk}\, dO_{11y}\ldots dO_{qky}\ldots dO_{11n}\ldots dO_{qkn} =$$

$$+\infty \qquad (12)$$

The likelihood function in equation (11) is the product of the pq probabilities of the observed choices. The value of the function is a maximum of 1 and a minimum of zero. Note that if all the legislators are voting correctly, that is, $\mathbf{P}_{\mathbf{ijc}} > .5$ (or $u_{ijc} > u_{ijb}$) for all i and j where “c” means “correct choice” and “b” means “incorrect choice”, then as $\beta \rightarrow +\infty$ clearly $\mathbf{L}(\, u_{ijc}\text{-}u_{ijb} \mid \mathbf{Y}\,) \rightarrow 1$. Conversely, if for at least one choice a legislator votes “incorrectly”, $u_{ijc} < u_{ijb}$ so that $\mathbf{P}_{\mathbf{ijc}} < .5$, then as $\beta \rightarrow +\infty$ clearly $\mathbf{L}(\, u_{ijc}\text{-}u_{ijb} \mid \mathbf{Y}\,) \rightarrow 0$. With voting error $\mathbf{L}(\, u_{ijy}\text{-}u_{ijn} \mid \mathbf{Y}\,)$ asymptotes very quickly to zero because $\Phi(u_{ijc}\text{-}u_{ijb})$ goes to zero very quickly as β increases.

Now consider the effect of the legislator and outcome coordinates. Suppose $|X_{ik}| \rightarrow +\infty$ then a simple inspection of equation (7) shows that:

$$\Phi\left[\beta\left\{e^{\left(-\frac{1}{2}\sum_{k=1}^{s} d_{ijkc}^{2}\right)} - e^{\left(-\frac{1}{2}\sum_{k=1}^{s} d_{ijkb}^{2}\right)}\right\}\right] \rightarrow \Phi\left[\beta\left\{e^{(-\infty)} - e^{(-\infty)}\right\}\right] = \Phi\left[\beta\left\{0\right\}\right] = .5$$

The portion of the likelihood function corresponding to legislator i is:

$$\prod_{j=1}^{q} \Phi(u_{ijc} - u_{ijb}) \qquad (13)$$

So that this converges to $.5^q$ as $|X_{ik}| \rightarrow +\infty$. This shows that $\mathbf{L}(u_{ijy}-u_{ijn} \mid \mathbf{Y})$ does not asymptote to zero along the dimensions corresponding to legislator coordinates so that the hypervolume underneath $\mathbf{L}(u_{ijy}-u_{ijn} \mid \mathbf{Y})$ is infinite.

The fact that $\mathbf{L}(u_{ijy}-u_{ijn} \mid \mathbf{Y})$ has an infinite hypervolume has no practical effect on a standard maximum likelihood analysis. This is so because at a great distance from the center of the space defined by the legislator and outcome points the likelihood function is a flat, featureless vista. That is, the maxima are towards the interior of the function and are easily found by conventional gradient and quasi-gradient methods. However, the use of simulation methods is inappropriate because $\mathbf{L}(u_{ijy}-u_{ijn} \mid \mathbf{Y})$ cannot be treated as if it were a probability distribution.

The Bayesian approach avoids the problem of infinite volume through the judicious choice of prior distributions that when multiplied through the likelihood function produce a distribution that is proportional to a probability distribution. In a standard Bayesian approach the prior distribution for a legislator ideal point is a multivariate normal distribution with variance-covariance matrix $\sigma^2\mathbf{I_s}$:

$$\xi(\mathbf{X_i}) = \frac{1}{(2\pi\sigma)^{\frac{s}{2}}} e^{-\frac{1}{2\sigma^2}(X_{i1}^2+X_{i2}^2+\ldots+X_{is}^2)} \tag{14}$$

Similarly, assume that the prior distributions for the outcome points are also multivariate normal distributions with variance-covariance matrices $\sigma^2\mathbf{I_s}$:

$$\xi(\mathrm{O_{jy}}) = \frac{1}{(2\pi\sigma)^{\frac{s}{2}}} e^{-\frac{1}{2\sigma^2}(O_{j1y}^2+O_{j2y}^2+\ldots+O_{jsy}^2)}$$

and (15)

$$\xi(\mathrm{O}_{\mathrm{jn}}) = \frac{1}{(2\pi\sigma)^{\frac{s}{2}}} e^{-\frac{1}{2\sigma^2}(O_{j1n}^2 + O_{j2n}^2 + ... + O_{jsn}^2)}$$

The posterior distribution for the NOMINATE model is:

$$\xi(\beta, \mathbf{O_y}, \mathbf{O_n}, \mathbf{X} \mid \mathbf{Y}) \propto \prod_{i=1}^{p}\prod_{j=1}^{q}\left\{f_{ij}(y_{ij} \mid u_{ijy} - u_{ijn})\xi(X_i)\xi(O_{jy})\xi(O_{jn})\right\} =$$

$$\prod_{i=1}^{p}\prod_{j=1}^{q}\left\{[\Phi(u_{ijy} - u_{ijn})]^{y_{ij}}[1-\Phi(u_{ijy} - u_{ijn})]^{(1-y_{ij})}\xi(X_i)\xi(O_{jy})\xi(O_{jn})\right\} =$$

$$\prod_{i=1}^{p}\prod_{j=1}^{q}\xi(X_i)\xi(O_{jy})\xi(O_{jn})\prod_{\tau=1}^{2}P_{ij\tau}^{C_{ij\tau}} \qquad (16)$$

By definition

$\xi(\beta, \mathbf{O_y}, \mathbf{O_n}, \mathbf{X} \mid \mathbf{Y}) \geq 0$, $0 < \beta < +\infty$, $-\infty < X_{ik}$, O_{jky} , $O_{jkn} < +\infty$

and the hypervolume underneath $\xi(\beta, \mathbf{O_y}, \mathbf{O_n}, \mathbf{X} \mid \mathbf{Y})$ is finite. Specifically, as the legislator ideal points and/or the outcome points go to $\pm\infty$ then $\xi(\beta, \mathbf{O_y}, \mathbf{O_n}, \mathbf{X} \mid \mathbf{Y})$ goes to zero. For example, as $|X_{ik}| \rightarrow +\infty$ then $\xi(\mathbf{X_i}) \rightarrow 0$ so that $\xi(\beta, \mathbf{O_y}, \mathbf{O_n}, \mathbf{X} \mid \mathbf{Y}) \rightarrow 0$.

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}...\int_{0}^{+\infty}\xi(\beta,\ \mathrm{O_y}, \mathrm{O_n}, \mathrm{X} \mid \mathrm{Y})d\beta dX_{11}...dX_{pk}dO_{11y}...dO_{qky}...dO_{11n}...dO_{qkn} = K < +\infty$$

Theoretically, simulation methods can be applied to the posterior distribution expressed in equation (16) but not to the likelihood distribution expressed in equation (11). The irony here is that if “believable” standard errors are to be obtained for the basic model either the parametric bootstrap (L&P) or the simulation approach are the best methods to do so because inverting the information matrix is problematic. I say “ironic” because one could favor the use of (16) over (11) on these practical grounds rather than the philosophical grounds of the “Bayesians” (non-“Frequentists”).

However, there is good reason to be skeptical of the Bayesian approach. What in fact are we modeling when we multiply the likelihood function by **$\xi(\mathbf{X_i})$?** Either we believe that legislators have ideal *points* or that "ideal points" are in fact *distributions* from which a decision maker makes a momentary psychological draw when a decision is made. This is indeed the model used by many Psychologists (see Poole, 2005, for a discussion of this literature). To multiply the likelihood function by $\xi(\mathbf{X_i}) \sim N(0,1)$ does not model this. Instead we would want to assume that

$$X_i \sim N(\mu_{X_i}, \sigma^2_{X_i} I_s) \qquad (17)$$

Where μ_{X_i} is the mean of the ideal point distribution with variance-covariance matrix $\sigma^2_{X_i} I_s$ and we would expect that $\sigma^2_{X_i}$ would be quite small.

Alternatively, we can regard the prior distribution $\xi(\mathbf{X_i}) \sim N(0,1)$ as a reflection of our *personal uncertainty* about $\mathbf{X_i}$**.** This subjectivist approach is more about us than modeling the behavior of legislator i. However, from another perspective multiplying the likelihood function by $\xi(\mathbf{X_i}) \sim N(0,1)$ is *neither* an expression of our personal uncertainty or a statement that legislator ideal points are distributions. Rather, we can simply state that it operates as a *constraint* on $\mathbf{X_i}$; that is, the $\mathbf{X_i}$ of relatively extreme legislators are not allowed to drift off into outer space.

Bottom line: in the computer code we should use the natural log of equation (16) (see below) and use classical methods to find the maxima for large problems and the parametric bootstrap to obtain the standard errors. For smaller problems we can experiment with incorporating the simulation methods to find the standard errors.

Before considering the logs of the likelihood and Posterior distributions, it is instructive to consider the implications of equation (17). In the Shepard-Ennis-Nofosky model of stimulus comparison people use a simple mental model to compare two stimuli. The distinguishing features of the stimuli are assumed to be represented by dimensions in a simple geometric model. The stimuli are positioned on the dimensions according to the levels of the attributes represented by the dimensions. People are assumed to perceive the stimuli correctly with some random error. When asked to perform a stimulus comparison, people draw a momentary psychological value from a very tight error distribution around the locations of each of the stimuli. Their judgment of similarity (the *response function*) is assumed to be an exponential function of the psychological distance between the two stimuli -- e^{-kd} -- where d is the distance between the two momentary psychological values expressed as points in psychological space, and $k > 0$ is a scaling constant (see the figure below).

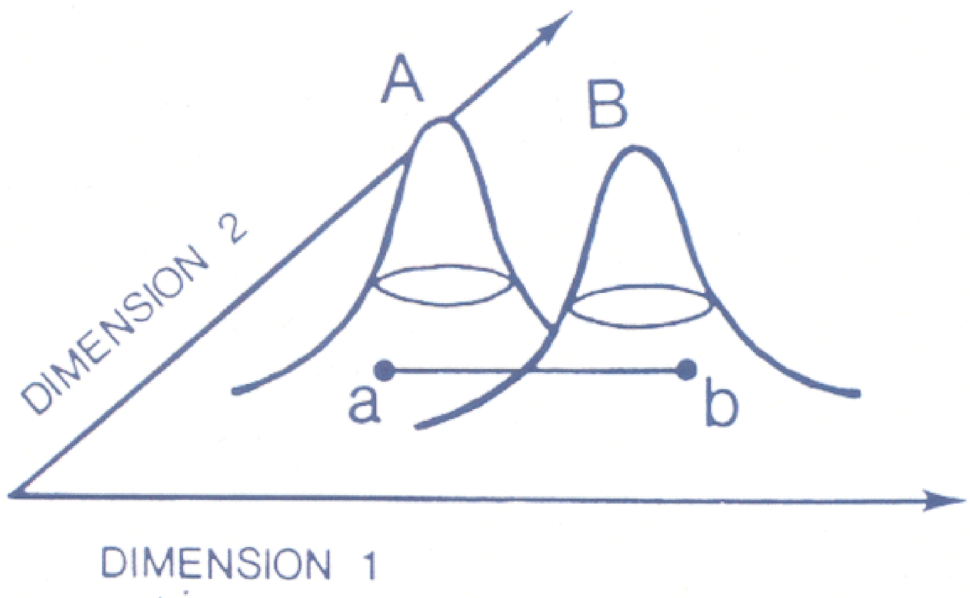


Figure 1. Exemplars A and B are represented as two-dimensional, Gaussian distributed psychological dispersions. The similarity between any two individual dispersions, *a* and *b*, would be an exponential decay function of their psychological distance: $s_{ab} = \exp(-d_{ab})$. The overall similarity between Exemplars A and B would reflect both the Gaussian noise and the exponential similarity function.

Suppose an individual is comparing her ideal point with a policy outcome under this model. That is, assume that:

$$O_j \sim N(\mu_{O_j}, \sigma^2_{O_j} I_s) \qquad (18)$$

Let the distance between the two draws be:

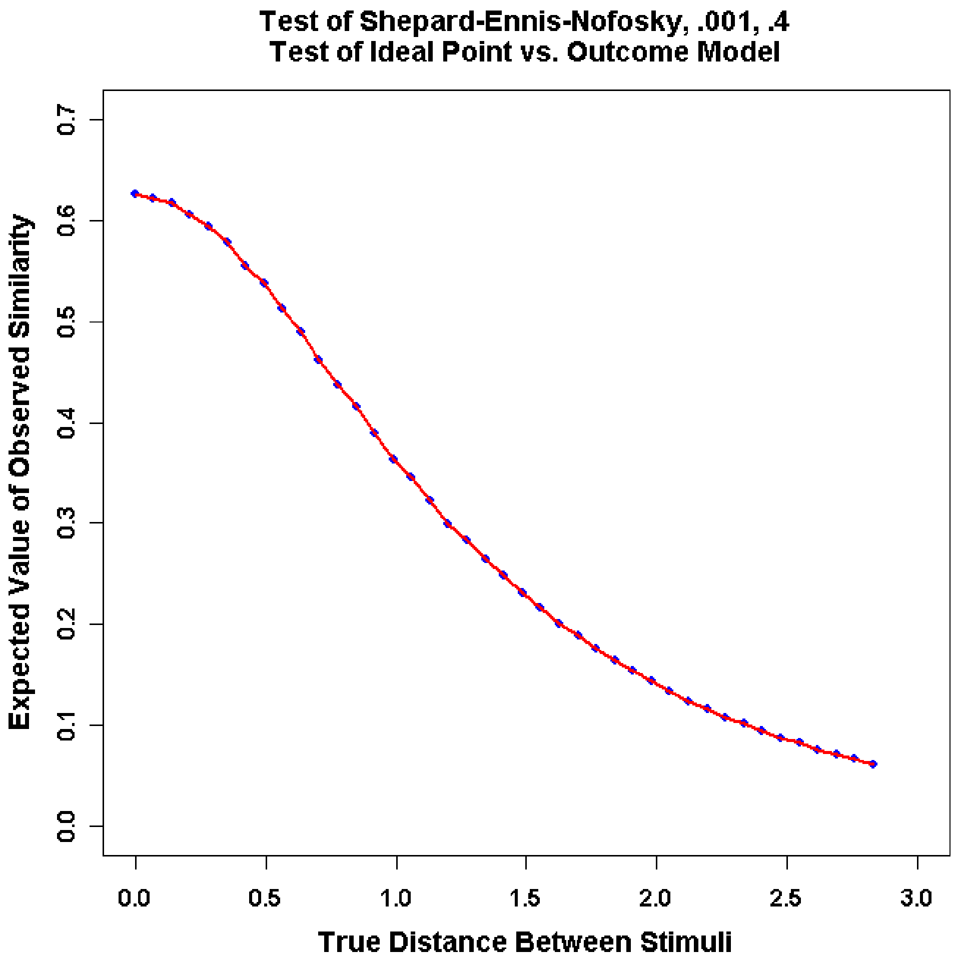
$$d_{ij} = \left[\sum_{k=1}^{s} (\tilde{X}_{ik} - \tilde{O}_{jk})^2 \right]^{\frac{1}{2}}$$

where $\tilde{X}_i$ and $\tilde{O}_j$ are the two points drawn from equations (17) and (18), respectively.

Hence the perceived similarity would be

$$S_{ij} = e^{-kd_{ij}}$$

The graph below shows the *expected value* of $e^{-kd_{ij}}$ (vertical axis) for 40 values of the true distance (horizontal axis) between μ_{X_i} and μ_{O_j} in two dimensions with k=1. The expected value was found by doing 10,000 draws from the two bivariate normal distributions with standard deviations of $\sigma_{X_i} = 0.001$ and $\sigma_{O_j} = 0.40$ respectively, computing the distance, exponentiating the result and taking the mean over the 10,000 draws.



Even though the legislator distribution is essentially a point the expected value of the response function is Gaussian. What this shows is that the Shepard-Ennis-Nofosky response function theory is basically equivalent to the basic NOMINATE random utility model shown in equation (1). That is, in the S-E-N model a legislator would draw a momentary psychological value from her ideal point distribution and values from the Yea and Nay outcome distributions. Hence if $e^{-kd_{ijy}} > e^{-kd_{ijn}}$ vote Yea and if $e^{-kd_{ijy}} < e^{-kd_{ijn}}$ vote Nay. Note that the error (randomness) in the model is embedded in d_{ijy} and d_{ijn} so that

$$\text{P(i votes Yea on j}^{\text{th}}\text{ roll call)} = P\left(e^{-kd_{ijy}} > e^{-kd_{ijn}}\right) = P\left(e^{-kd_{ijy}} - e^{-kd_{ijn}} > 0\right)$$

This would be an interesting model to try to estimate. I am unclear what a Bayesian approach would be here other than the prior assumption that all the distributions are multivariate normal but this is more like a classical modeling assumption. Equation (17) could not be used as a *prior* distribution because μ_{X_i} and $\sigma^2_{X_i}$ are parameters we would want to estimate. Furthermore, if we assumed we knew μ_{X_i} as we do when we use vague priors then the whole enterprise would be pointless.

Returning to the main model the natural log of the likelihood function in equation (11) is:

$$\Xi = \ln\{L(u_{ijy}\text{-}u_{ijn} \mid \mathbf{Y})\} = \sum_{\mathbf{i=1}}^{p}\sum_{\mathbf{j=1}}^{q}\sum_{\boldsymbol{\tau=1}}^{\mathbf{2}} C_{ij\tau} \ln P_{ij\tau} = \sum_{i=1}^{p}\sum_{j=1}^{q} \ln\Phi\left[\beta\psi_{ijc}\right] \qquad (19)$$

Where

$$\psi_{ijc} = e^{\left(-\frac{1}{2}\sum_{k=1}^{s} d^2_{ijkc}\right)} - e^{\left(-\frac{1}{2}\sum_{k=1}^{s} d^2_{ijkb}\right)}$$

and the subscript "c" stands for the observed choice and "b" for the alternative not chosen. In terms of estimation it is convenient to write the roll call outcome coordinates as functions of the midpoint and half the distance between the outcomes:

$$O_{jkc} = Z_{jkm} - \delta_{jkc}$$

$$O_{jkb} = Z_{jkm} + \delta_{jkc}$$

where

$$Z_{jkm} = \frac{\left(O_{jky} + O_{jkn}\right)}{2} \quad \text{and} \quad \delta_{jkc} = \frac{\left(O_{jkb} - O_{jkc}\right)}{2}$$

However, for the time being I will stick with O_{jkc} and O_{jkb} because it aids in the interpretation of the derivatives.

The first derivatives of equation (17) are below. It is helpful when examining them to recall the chain rule for the normal distribution and its integral:

$$\frac{\partial \Phi(u)}{\partial u} = f(u)du \quad \text{and} \quad \frac{\partial f(u)}{\partial u} = f(u)(-u)du$$

and the basic rule for the derivative of a ratio: $\partial\left[\frac{u}{v}\right] = \frac{vdu - udv}{v^2}$.

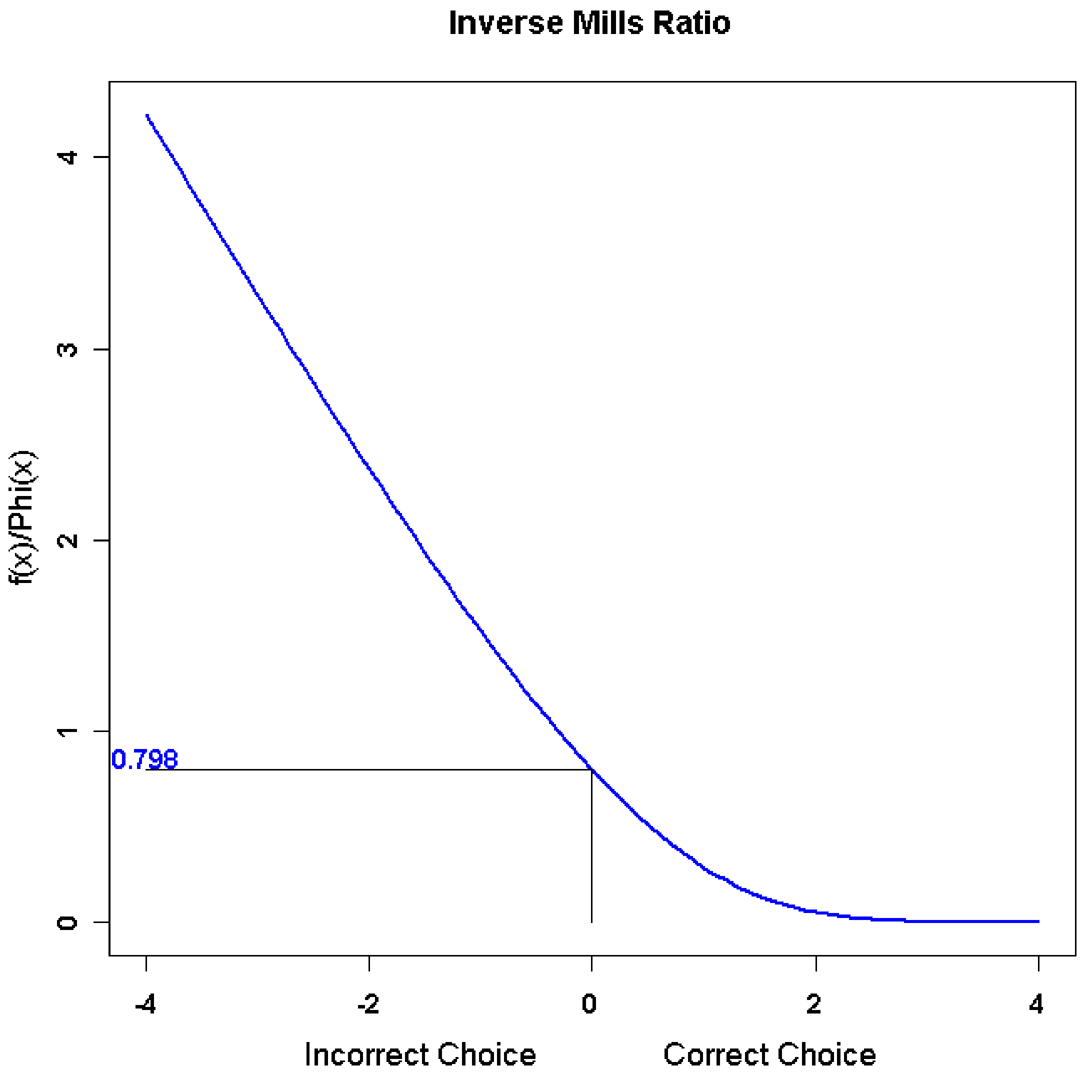
$$\frac{\partial \Xi}{\partial \beta} = \sum_{i=1}^{p}\sum_{j=1}^{q}\frac{f\left(\beta\psi_{ijc}\right)}{\Phi\left(\beta\psi_{ijc}\right)}\psi_{ijc} \qquad (20)$$

$$\frac{\partial \Xi}{\partial X_{ik}} = \beta\sum_{j=1}^{q}\left\{\frac{f\left(\beta\psi_{ijc}\right)}{\Phi\left(\beta\psi_{ijc}\right)}\left[-e^{\left(-\frac{1}{2}\sum_{k=1}^{s}d_{ijkc}^2\right)}\left(X_{ik} - O_{jkc}\right) + e^{\left(-\frac{1}{2}\sum_{k=1}^{s}d_{ijkb}^2\right)}\left(X_{ik} - O_{jkb}\right)\right]\right\} \qquad (21)$$

$$\frac{\partial \Xi}{\partial O_{jkc}} = \beta\sum_{i=1}^{p}\left\{\frac{f\left(\beta\psi_{ijc}\right)}{\Phi\left(\beta\psi_{ijc}\right)}\left[e^{\left(-\frac{1}{2}\sum_{k=1}^{s}d_{ijkc}^2\right)}\left(X_{ik} - O_{jkc}\right)\right]\right\} \qquad (22)$$

$$\frac{\partial \Xi}{\partial O_{jkb}} = \beta\sum_{i=1}^{p}\left\{\frac{f\left(\beta\psi_{ijc}\right)}{\Phi\left(\beta\psi_{ijc}\right)}\left[-e^{\left(-\frac{1}{2}\sum_{k=1}^{s}d_{ijkb}^2\right)}\left(X_{ik} - O_{jkb}\right)\right]\right\} \qquad (23)$$

The derivatives shown above all have a similar form in that they all contain the inverse Mill's Ratio, $\frac{f\left(\beta\psi_{ijc}\right)}{\Phi\left(\beta\psi_{ijc}\right)}$, which applies a larger positive weight to the incorrect choices relative to the correct choices (see graph below):



The effect of the inverse Mills Ratio is best illustrated by examining the first and second derivatives for β. The second derivative for β is:

$$\frac{\partial^2 \Xi}{\partial \beta^2} = -\sum_{i=1}^{p}\sum_{j=1}^{q} \psi_{ijc}^2 \frac{f(\beta\psi_{ijc})}{\Phi(\beta\psi_{ijc})}\left[\beta\psi_{ijc} + \frac{f(\beta\psi_{ijc})}{\Phi(\beta\psi_{ijc})}\right] \qquad (24)$$

Now, setting the first derivative for β equal to zero and rearranging:

$$\sum_{\text{correct}} \sum \frac{f(\beta\psi_{ijc})}{\Phi(\beta\psi_{ijc})}\psi_{ijc} = -\sum_{\text{incorrect}} \sum \frac{f(\beta\psi_{ijc})}{\Phi(\beta\psi_{ijc})}\psi_{ijc} \qquad (25)$$

Rearranging the second derivative in the same fashion produces:

$$\frac{\partial^2 \Xi}{\partial \beta^2} = -\sum_{i=1}^{p}\sum_{j=1}^{q}\psi_{ijc}^2\left[\frac{f(\beta\psi_{ijc})}{\Phi(\beta\psi_{ijc})}\right]^2 - \beta\sum_{\text{correct}}\sum\psi_{ijc}^3\frac{f(\beta\psi_{ijc})}{\Phi(\beta\psi_{ijc})} - \beta\sum_{\text{incorrect}}\sum\psi_{ijc}^3\frac{f(\beta\psi_{ijc})}{\Phi(\beta\psi_{ijc})} \qquad (26)$$

For incorrect choices $\psi_{ijc} < 0$ and the corresponding $\frac{f(\beta\psi_{ijc})}{\Phi(\beta\psi_{ijc})}$ is larger than for correct choices, $\psi_{ijc} > 0$. In addition, note that $\sum_{i=1}^{p}\sum_{j=1}^{q}\psi_{ijc}^2\left[\frac{f(\beta\psi_{ijc})}{\Phi(\beta\psi_{ijc})}\right]^2 > 0$,

$\beta\sum_{\text{correct}}\sum\psi_{ijc}^3\frac{f(\beta\psi_{ijc})}{\Phi(\beta\psi_{ijc})} > 0$, and $\beta\sum_{\text{incorrect}}\sum\psi_{ijc}^3\frac{f(\beta\psi_{ijc})}{\Phi(\beta\psi_{ijc})} < 0$.

Suppose that legislators simply flip fair coins so that the roll call votes are clustered around 50-50. As Howard and I determined 20 years ago NOMINATE will scale this data and achieve a correct classification about equal to:

$$\text{Correct Classification} = \frac{\sum_{j=1}^{q}\left(\text{Majority Side of Roll Call}\right)_j}{\text{Total Votes Cast}} * \mathbf{100}$$

Note that this number will be *slightly above* 50 percent. In this instance β will be quite small because of the asymmetry of the inverse Mills Ratio. With β close to zero the $\frac{f(\beta\psi_{ijc})}{\Phi(\beta\psi_{ijc})}$ terms will all be close to approximately .798 (see figure above) so that equation (25) will be satisfied. Furthermore, $\frac{\partial^2 \Xi}{\partial \beta^2} < 0$ because the first term of equation

(26) will dominate the third term because of the multiplication of the third term by β so that the inflection point will be a maximum.

In contrast, consider the effect of a very high rate of correct classification. In this case β will become large. Because there are few classification errors, for equation (25) to hold β must be large to drive up the magnitude of the $\frac{f\left(\beta\psi_{ijc}\right)}{\Phi\left(\beta\psi_{ijc}\right)}$ terms corresponding to the classification errors. This in turn drives down the $\frac{f\left(\beta\psi_{ijc}\right)}{\Phi\left(\beta\psi_{ijc}\right)}$ terms corresponding to the correct classifications so that equation (25) has a solution. Also $\frac{\partial^2\Xi}{\partial\beta^2} < 0$ because with a small number of errors the second and third terms of equation (26) will be approximately equal. To see this note that, ignoring the multiplication by β, the second term is the same as the left hand side of equation (25) only ψ_{ijc} has been replaced by ψ_{ijc}^3. Similarly, the third term of equation (25) is the same as the right side of equation (26) only ψ_{ijc} has been replaced by ψ_{ijc}^3; that is, each term on both sides is weighted by the square of its corresponding ψ_{ijc} term and $\psi_{ijc}^2 < 1$. In other words, suppose $\frac{\partial\Xi}{\partial\beta} = 0$ then

$$\sum_{correct}\sum\frac{f\left(\beta\psi_{ijc}\right)}{\Phi\left(\beta\psi_{ijc}\right)}\psi_{ijc}^3 \approx -\sum_{incorrect}\sum\frac{f\left(\beta\psi_{ijc}\right)}{\Phi\left(\beta\psi_{ijc}\right)}\psi_{ijc}^3$$

so that the second and third terms in equation (26) roughly cancel out and $\frac{\partial^2\Xi}{\partial\beta^2} < 0$ because of the first term.

Before turning to the natural log of the Posterior distribution in equation (16) note that the first derivative for the legislator coordinate can be rearranged into an implicit equation:

$$X_{ik} = \frac{\sum_{j=1}^{q}\left\{\frac{f\left(\beta\psi_{ijc}\right)}{\Phi\left(\beta\psi_{ijc}\right)}\left[e^{\left(-\frac{1}{2}\sum_{k=1}^{s}d_{ijkc}^{2}\right)}O_{jkc} - e^{\left(-\frac{1}{2}\sum_{k=1}^{s}d_{ijkb}^{2}\right)}O_{jkb}\right]\right\}}{\left[\sum_{j=1}^{q}\frac{f\left(\beta\psi_{ijc}\right)}{\Phi\left(\beta\psi_{ijc}\right)}\psi_{ijc}\right]} = g(X_{ik}) \qquad (27)$$

Implicit equations are quite useful in that you can plug in a trial value into $g(X_{ik})$, get a result, plug in a modified value depending on the result, etc., and iterate into a solution. The downside here is that equation (27) is *very* non-linear but it only involves looping over the roll call votes cast by the i^{th} legislator. I will discuss this in more detail below.

The natural log of the Posterior Distribution shown in equation (16) is:

$$\begin{aligned}\Theta = \ln\left(\prod_{i=1}^{p}\prod_{j=1}^{q}\xi(X_i)\xi(O_{jy})\xi(O_{jn})\prod_{\tau=1}^{2}P_{ij\tau}^{C_{ij\tau}}\right) = \\ \sum_{i=1}^{p}\sum_{j=1}^{q}\ln\Phi\left[\beta\psi_{ijc}\right] + \sum_{i=1}^{p}\ln\xi(X_i) + \sum_{j=1}^{q}\ln\xi(O_{jy}) + \sum_{j=1}^{q}\ln\xi(O_{jn})\end{aligned} \qquad (28)$$

This is the same as the natural log of the Likelihood function shown in equation (19) with the addition of the natural logs of the prior distributions. Note that I do not have a prior distribution for β.

$$\frac{\partial\Theta}{\partial\beta} = \frac{\partial\Xi}{\partial\beta} = \sum_{i=1}^{p}\sum_{j=1}^{q}\frac{f\left(\beta\psi_{ijc}\right)}{\Phi\left(\beta\psi_{ijc}\right)}\psi_{ijc} \qquad (29)$$

$$\frac{\partial\Xi}{\partial X_{ik}} = \beta\sum_{j=1}^{q}\left\{\frac{f\left(\beta\psi_{ijc}\right)}{\Phi\left(\beta\psi_{ijc}\right)}\left[-e^{\left(-\frac{1}{2}\sum_{k=1}^{s}d_{ijkc}^{2}\right)}\left(X_{ik} - O_{jkc}\right) + e^{\left(-\frac{1}{2}\sum_{k=1}^{s}d_{ijkb}^{2}\right)}\left(X_{ik} - O_{jkb}\right)\right]\right\} - \frac{X_{ik}}{\sigma^2} \qquad (30)$$

$$\frac{\partial \Xi}{\partial O_{jkc}} = \beta \sum_{i=1}^{p} \left\{ \frac{f\left(\beta \psi_{ijc}\right)}{\Phi\left(\beta \psi_{ijc}\right)} \left[e^{\left(-\frac{1}{2}\sum_{k=1}^{s} d_{ijkc}^2\right)} \left(X_{ik} - O_{jkc}\right) \right] \right\} - \frac{O_{jkc}}{\sigma^2} \qquad (31)$$

$$\frac{\partial \Xi}{\partial O_{jkb}} = \beta \sum_{i=1}^{p} \left\{ \frac{f\left(\beta \psi_{ijc}\right)}{\Phi\left(\beta \psi_{ijc}\right)} \left[-e^{\left(-\frac{1}{2}\sum_{k=1}^{s} d_{ijkb}^2\right)} \left(X_{ik} - O_{jkb}\right) \right] \right\} - \frac{O_{jkb}}{\sigma^2} \qquad (32)$$

These derivatives are the same as those shown in equations (21), (22), and (23) with the addition of the final terms from differentiating the normal prior distributions -- $-\frac{X_{ik}}{\sigma^2}, -\frac{Z_{jk}}{\sigma^2}$, and $-\frac{\delta_{jk}}{\sigma^2}$ -- respectively. I do not subscript the variance terms – the σ^2's – deliberately. I believe we will want to experimentally determine the best values for these. For example, the variance for the prior of a midpoint probably should be smaller than the variance for the prior of a legislator point.

The potential role these variance terms may play is best illustrated by looking at the implicit equation for a legislator coordinate:

$$X_{ik} = \frac{\sum_{j=1}^{q} \left\{ \frac{f\left(\beta \psi_{ijc}\right)}{\Phi\left(\beta \psi_{ijc}\right)} \left[e^{\left(-\frac{1}{2}\sum_{k=1}^{s} d_{ijkc}^2\right)} O_{jkc} - e^{\left(-\frac{1}{2}\sum_{k=1}^{s} d_{ijkb}^2\right)} O_{jkb} \right] \right\}}{\left[\sum_{j=1}^{q} \frac{f\left(\beta \psi_{ijc}\right)}{\Phi\left(\beta \psi_{ijc}\right)} \psi_{ijc} \right] + \frac{1}{\sigma^2}} = g^*(X_{ik}) \qquad (33)$$

Equation (33) is the same as equation (27) except for the $\frac{1}{\sigma^2}$ in the denominator. Now if the prior distribution was *uninformative*, that is, if $\sigma^2 = 1000$ or something similar, then the prior distribution has no impact on the derivatives and hence no impact on any gradient style algorithm used to find the maxima of the parameters. However, this is an unsatisfactory result for our purposes because we need to use the priors to constrain the distribution of the estimated parameters.

In this regard, note that $\sum_{j=1}^{q} \frac{f\left(\beta\psi_{ijc}\right)}{\Phi\left(\beta\psi_{ijc}\right)} \psi_{ijc} < q$. Furthermore, suppose the legislator makes no voting errors then all the $\frac{f\left(\beta\psi_{ijc}\right)}{\Phi\left(\beta\psi_{ijc}\right)}$ terms are less than 0.8. Indeed, if the average probability is around 0.7 – Φ(.5)=0.69 -- then $\frac{f(.5)}{\Phi(.5)} \approx .5$ and if β=2.5 then the $\psi_{ijc} \approx 0.2$ so that $\sum_{j=1}^{q} \frac{f\left(\beta\psi_{ijc}\right)}{\Phi\left(\beta\psi_{ijc}\right)} \psi_{ijc} \approx 0.10q$. Suppose the legislator was to the exterior of all the midpoints, that is $X_{ik} > Z_{jkm}\ \forall\ j$ so that $O_{jkc} > O_{jkb}\ \forall\ j$. This will tend to produce a "sag". To see this note that – ignoring the $\frac{\mathbf{1}}{\boldsymbol{\sigma}^2}$ in the denominator – for each j the numerator term is larger than its corresponding denominator term:

$$\left[e^{\left(-\frac{1}{2}\sum_{k=1}^{s} d_{ijkc}^2\right)} O_{jkc} - e^{\left(-\frac{1}{2}\sum_{k=1}^{s} d_{ijkb}^2\right)} O_{jkb} \right] > \left[e^{\left(-\frac{1}{2}\sum_{k=1}^{s} d_{ijkc}^2\right)} - e^{\left(-\frac{1}{2}\sum_{k=1}^{s} d_{ijkb}^2\right)} \right] = \boldsymbol{\psi}_{ijc} > \mathbf{0},\ \forall\ j$$

Because $e^{\left(-\frac{1}{2}\sum_{k=1}^{s} d_{ijkc}^2\right)} > e^{\left(-\frac{1}{2}\sum_{k=1}^{s} d_{ijkb}^2\right)}$ and $O_{jkc} > O_{jkb}$. So with respect to equation (27) -- the implicit equation without $\frac{\mathbf{1}}{\boldsymbol{\sigma}^2}$ -- $g(X_{ik}) > 1$. Now, let X_{ik} increase in positive magnitude. The $\frac{f\left(\beta\psi_{ijc}\right)}{\Phi\left(\beta\psi_{ijc}\right)}$ terms become smaller and smaller and approach zero; $e^{\left(-\frac{1}{2}\sum_{k=1}^{s} d_{ijkc}^2\right)} \approx e^{\left(-\frac{1}{2}\sum_{k=1}^{s} d_{ijkb}^2\right)}$ and $g(X_{ik}) \approx 1$. In other words, the quasi-concave form of the utility function keeps X_{ik} from "blowing up" because there is an inflection point some distance from the "sensible boundary" of the voting space (Poole, 2005, p. 97). This is a serious problem because the legislator point for a "perfect" legislator drifts out from the other legislators and creates a sizeable "sag" that lacks face validity in the analysis of real-world legislatures.

Note that perfect legislators in the *interior* of the space are not a problem. This is due to the simple fact that they are *in the midst* of midpoints and if X_{ik} crosses over any midpoint then the legislator is no longer perfect! The denominator term $\sum_{j=1}^{q} \frac{f\left(\beta\psi_{ijc}\right)}{\Phi\left(\beta\psi_{ijc}\right)} \psi_{ijc}$ will be positive but not as small as that for a perfect legislator near the rim of the space because the $e^{\left(-\frac{1}{2}\sum_{k=1}^{s} d_{ijkb}^{2}\right)}$ terms will not be as small relative to the $e^{\left(-\frac{1}{2}\sum_{k=1}^{s} d_{ijkc}^{2}\right)}$ terms and the numerator can be positive or negative because the O_{jkc} can be to the right or to the left of the legislator coordinate.

The $\frac{\mathbf{1}}{\boldsymbol{\sigma}^{2}}$ in the denominator of equation (33) will act as a break on the outward drift of a "perfect" legislator near the rim of the space for the very simple mathematical reason that it is fixed in magnitude vis a vis the $\sum_{j=1}^{q} \frac{f\left(\beta\psi_{ijc}\right)}{\Phi\left(\beta\psi_{ijc}\right)} \psi_{ijc}$ term which can become quite small. What the value of σ^2 should be will have to be determined by experimentation during the first phase of development of the new algorithm in C. My intuition is that $\sigma^2=1$ will probably do the trick. Note that this assumption expresses our *belief* that legislators *tend* to be within a multivariate normal distribution with variance-covariance matrix $\mathbf{I_s}$**.** There may be some interaction between this σ^2 and those for the assumed distributions of the roll call parameters. We shall see.